

DR ProDom; PD000425; TF Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00659; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 SQ SEQUENCE 371 AA; 40018 MW; 234645747EALF716 CRC64;

Alignment Scores:
 Pred. No.: 6,78e-13 Length: 371
 Score: 402.00 Matches: 124
 Percent Similarity: 43.14% Conservative: 52
 Best Local Similarity: 30.39% Mismatches: 126
 Query Match: 18.36% Indels: 106
 DB: 13 Gaps: 17

US-10-087-080-31 (1-1209) x Q9DEN4 (1-371)

QY 37 GGGGACAGAGCGGCGTACCTGGAGGC-----GCGGGCGGACGAGCGGCG 87
 Db 5 GlycerGlySerAlaSerMetSerGlyThrValLeuSerAlaAspAlaAsp 24
 QY 88 TCCCGCTGTCCGCGGCGGAGAGCACTCCCTGGGCTCAGATGGGACTCCGCGGCCAAG 147
 Db 25 IleAspValValGlyGlyGlyAspGluAlaLeuAspLysAspSerGluCys----- 41
 QY 148 CCGTCCCGGCGGCGGCGGCGGAGATACGAGGGCGACGCGAAAGAGTGGGAGGC 207
 Db 42 ---GluSerThrAlaGlyHisThrAspGluValGlyGlu-----LeuGlyGly 56
 QY 208 GGGCGGCGGCGGAGGAGGCGATCCCGGCGAGCTGCTGCAGCGGTGGTGGCGGAGGC 267
 Db 57 -----LysGluLeuProArgSerProSer-----Gly 65
 QY 268 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 327
 Db 66 SerGlyThrGluAlaGlyGlyLysGlyGlySerGlnGlnGlnGlnGlnGlyLeuGln 85
 QY 328 AGCAAGCATATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 387
 Db 86 AsnLysProLysAsnSerLeuValLysProProTySerTyLysAlaLeuLeuThrMet 105
 QY 388 GCCATCCGCGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 447
 Db 106 SerIleLeuGlnSerProGlnLysLysLeuThrLeuSerGlyLysCysGluPheIleSer 125
 QY 448 GGCAGTTCCTCTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 507
 Db 126 AsnArgPheProTyTyTrArgGlyLysPheProAlaTrpGlnAsnSerIleArgHisAsn 145
 QY 508 CTTTCGTCAACGACTCTCTGTCAGGTGCTGCGGACCCCTCGCGGCGGCGGCGGCGGCGG 567
 Db 146 LeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---GlyLys 164
 QY 568 GACNACTACTGGATGCTCAACCCCAACAGGAGTACACCTTCGCCGAGGGTCTTCGCG 627
 Db 165 GlyAsnTyTrpThrLeuAspProGlnSerGluAspMetPheAspAsnGlySerPheLeu 184
 QY 628 CGCGCGGCGCAAGCCCTCAGCCACCGCGCG----- 657
 Db 185 ArgArgArgLysArgPheLysArgGlnGlnAspSerLeuArgGluThrAlaLeu 204
 QY 658 -----CCGTCCCGCGCGCC-----GGGCTG 678
 Db 205 MetMetGlnSerPheGlyAlaTyTrSerLeuAlaSerProTyGlyArgHisTyGlyLeu 224
 QY 679 CGGCGCGGAG-----GAGCGCGCGGCGGCTCCCGCGCGGCGC--- 711
 Db 225 HisProAlaAlaTyThrHisProAlaAlaLeuGlnTyProTyTrIleProProValGly 244
 QY 712 -----CG 753
 Db 245 HisMetLeuProProAlaValProLeuLeuProSerSerGluLeuThrArgLysAlaPhe 264
 QY 754 -----TCGCGCGCGCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 801

Db 265 SerSerGlnLeuSerProSerLeuGlnLeuGlnLeuSerSerLeuSerSerThrAlaAla 284
 QY 802 AGC-----TCCTTCGCGCATCGACGACGATCCTCGCG 831
 Db 285 SerIleLeuLysSerGluProSerArgProSerPheSerIleGluAsnIleGly 304
 QY 832 AAGCCCTTCCCGAGCGCTCGCTCAGGACACAGCCCGCGGACGCGCTTCACTGGGCG 891
 Db 305 ValSerAlaAlaSer-----SerAlaAlaProHisThrPheLeuArg----- 318
 QY 892 GCGGCGCGCTCGCGCGCTCCCGCGGTTCCTCCCGCGCTCTCCCGCGGCGCGCTGCAGG 951
 Db 319 -----ProProValThrValGlnSerAlaLeuMet----- 328
 QY 952 GCCCTGCTCGCGCTCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 1011
 Db 329 -----SerHisGlnProLeuAlaLeuSerArgSerThr 339
 QY 1012 GCGGAGGTGCCACG-----ACGCGCGCGCGCTCTCTGCTTGA-----CCT 1053
 Db 340 AlaAlaIleGlyProIleLeuSerValProThrAsnLeuIleSerGlyGlnPheLeuPro 359
 QY 1054 CTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1077
 Db 360 ThrAlaAlaAlaAlaValAlaLys 367

RESULT 15
 Q96D28 PRELIMINARY; PRT; 330 AA.

AC Q96D28; (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 25, Last annotation update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Hypothetical protein (Forkhead-like 18) (Drosophila).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kalnine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
 RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
 RA Phelan M., Farmer A.;
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC013408; AAH13408.1; -;
 DR EMBL; BT009768; AAP88770.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; E:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00659; FORK_HEAD_3; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 330 AA; 35434 MW; BBF10321A8DC1744 CRC64;

Alignment Scores:
 Pred. No.: 1.64e-12 Length: 330
 Score: 394.50 Matches: 120
 Percent Similarity: 40.90% Conservative: 26
 Best Local Similarity: 33.61% Mismatches: 110
 Query Match: 18.02% Indels: 101
 DB: 4 Gaps: 14

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 04:45:19 ; Search time 4126 Seconds
(without alignments)
4222.952 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVPRAAHGKQGSDDL.....AALVRRFGPHLSYPVETLLA 402

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-O=/cgn2.1/USPTO.spool.p/US10087080/runat_29032004.114946.24387/app_query.fasta_1.583
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=200 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080@cgn.1.1.4958 -runat_29032004.114946.24387 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_othr.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rtd.*
36: em_hgt_ham.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_othr.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2123	100.0	1448	9	AF153341 Homo sapi
2	2093.5	98.6	62761	9	AL499606 Human DNA
3	2081.5	98.0	2322	9	BC053850 Homo sapi
4	2081.5	98.0	4109	9	AF225950 Homo sapi
5	1597	75.2	1513	10	AF154426 Mus muscu
6	1597	75.2	2046	10	BC047155 Mus muscu
7	1597	75.2	187709	10	AL589738 Mouse DNA
8	1589	74.8	4763	10	AF010405 Mus muscu
9	1534	72.3	1576	10	AF153193 Rattus ra
10	1534	72.3	195184	2	AC119497 Rattus no
11	1412.5	66.5	1760	10	L13201 Rattus norv
12	1372	64.6	182152	2	AC067929 Homo sapi
13	1159	54.6	178168	2	AC012350 Homo sapi
14	956.5	45.1	74969	2	AC016269 Homo sapi
15	685.5	32.3	62848	5	BX510653 Zebrafish
16	561.5	26.4	2222	3	AK112352 Ciona int
17	522	24.6	20236	5	GGU47276 Gallus gall
18	494.5	23.3	106239	9	AL607122 Human DNA
19	494.5	23.3	153604	2	AC022754 Homo sapi
20	493.5	23.2	4259	9	AF042832 Homo sapi
21	488.5	23.0	2272	9	HSU59832 Human trans
22	488.5	23.0	5181	6	AX336763 Sequence
23	487.5	23.0	1593	10	AF023915 Mus muscu
24	487.5	23.0	201142	10	AL670035 Mouse DNA
25	487.5	22.9	1708	5	GGU37272 Gallus gall
26	486	22.8	2437	10	MUSBF2TF
27	484.5	22.6	173073	9	AC099522 Homo sapi
28	479.5	22.6	182835	10	AC121839 Mus muscu
29	479.5	22.6	182835	10	AF142647
30	473.5	22.3	7005	10	AR102796 Sequence
31	468.5	22.1	3946	6	AR140209 Sequence
32	468.5	22.1	3946	6	BD226150 Glaucoma
33	468.5	22.1	3946	9	AF048693 Homo sapi
34	468.5	22.1	3946	9	AF048693 Homo sapi
35	468.5	22.1	104729	9	HS118818 Human DNA
36	467.5	22.0	3582	9	HSFREA1
37	467.5	22.0	120429	9	HS668J24
38	467	22.0	1465	5	GGU37274
39	466.5	22.0	2187	6	AX578019 Sequence
40	466.5	22.0	2187	9	HSU13230 Rattus no
41	465.5	21.9	209010	2	AC123435
42	462.5	21.8	1673	9	HST1F2
43	462	21.8	2011	9	AF275722 Homo sapi
44	462	21.8	142728	9	HSU792G4 Human DNA
45	460	21.7	217762	10	AL806523 Mouse DNA

ALIGNMENTS

RESULT 1

AF153341 1448 bp DNA linear PRI 24-AUG-2001

LOCUS Homo sapiens winged helix/forkhead transcription factor (HFH1)
gene, complete cds.

ACCESSION AF153341
VERSION AF153341
KEYWORDS AF153341.1 GI:8489092
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1448)
HONG,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
Justice,M.J. and Chakravarti,A.
The winged helix/forkhead transcription factor Foxq1 regulates
differentiation of hair in satin mice
Genes 29 (4), 163-171 (2001)
JOURNAL 21207067
MEDLINE 11309849
PUBMED 21207067
REFERENCE 2 (bases 1 to 1448)
HONG,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
Direct Submission
TITLE Submitted (21-MAY-1999) Genetics, Case Western Reserve University,
10900 Euclid Avenue, Cleveland, OH 44106-4955, USA
JOURNAL Location/Qualifiers
FEATURES
1..1448
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p25"
233..>1441
/gene="HFH1"
233..>1441
/gene="HFH1"
/product="winged helix/forkhead transcription factor"
233..1441
/gene="HFH1"
/codon_start=1
/product="winged helix/forkhead transcription factor"
/protein_id="AAF75596.1"
/db_xref="GI:8489093"
/translation="MKLEVPVRAAHGDKGSLDEAGSGSDAPSLAAGDSDLSGDG
DCAKPSAGAGARDTQGDQSQSGGGGAEAIAPAAAAVAEGAEAGAGAGGAGGA
GSGEAKNSKPYTRPRKPFYIYALIAWAIRDSAGRLTAEINELMGKFPFRGGYT
GWRNVNHNLDSCFVKVLRDSPRMKONVWNLNSETTFADGVFRRKRLRSH
APVPAQLRDEEAPGLPAPAPAPAPSPMRSPAQEERASPAQKFSFSAIDSL
RKPFRRRLRDATPGTTLQNGAAPCPPLPAPFPALLPAPCGRALLPCAYGAGGPAFIG
ARAEVPTPTAPULLALPLPAAPAKPLRGPAAAGAHLYCPLRLPAAIQALVRRPGRPH
LSYPTVETLLA"

ORIGIN

Alignment Scores:
Pred. No.: 3,53e-50 Length: 1448
Score: 2123.00 Matches: 402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-087-080-32 (1-402) x AF153341 (1-1448)

Qy 1 MetHysLeuGIuValpHeValProArgAlaAlaHisGlyAspIysGlnGlySerAspLeu 20
Db 233 ATGAAGTTGAGGTGTTCTGCTCCCTCGCGCGGCCACCGGGACAAAGCAGGCGAGTGCACCTG 292
Qy 21 GluCluValAcGlyGlySerAspAlaProSerProLeuSerIlaaIaGlyAspAspSerLeu 40
Db 293 GAGGCGCGCGCGGAGCGAGCGCGCCCTCCCTGTCGCGCGCGGGAGGACGACTCCCTG 352
Qy 41 GlySerAspGlyAspCysAlaAlaIaIaProSerXlaGlyGlyValaIaArgAspThrGln 60
Db 353 GGCTCAGATGGGACTCGCGCGGCGCAAGCGCTCGCGGCGCGCGCGCGGCGGACGATACGAC 412

RESULT 2
LOCUS AL499606 62761 bp DNA linear PRI 19-DEC-2001
DEFINITION Human DNA sequence from clone RP11-13J16 on chromosome 6, complete sequence.
ACCESSION AL499606
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 62761)
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hunquerry@sanger.ac.uk
COMMENT On Dec 20, 2001 this sequence version replaced gi:17902904. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-13J16 is from the library RPI1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
FEATURES
source
1..62761
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="RPI1-11.1"
/clone="RP11-13J16"
17171..17234
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
misc_feature
17634..17641
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
misc_feature
21165..21540
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Assembly consistent with restriction digest data."
misc_feature
21466
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 430bp by restriction digest data."
misc_feature
23824..24045
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
misc_feature
60092
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 120bp by restriction digest data."
ORIGIN
Alignment Scores:
Pred. No.: 4,1e-48 Length: 62761
Score: 2093.50 Matches: 398
Percent Similarity: 99.01% Conservative: 1
Best Local Similarity: 98.76% Mismatches: 3
Query Match: 98.61% Indels: 1
DB: 9 Gaps: 1
US-10-087-080-32 (1-402) x AL499606 (1-62761)
QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db ATGAAGTTGGAGGTGTTCCTCGCGCGGCCCAAGCAGGCGAGTACCTG 17004
QY 21 GluGlyAlaGlyCysAspAlaProSerProLeuSerAlaAlaGlyAspAspSer 40
Db GAGGGCGCGGGCGGCGAGCGCGCTCCCGCTGTCGGCGCGGCGGAGCAGCTCCCTG 17064
QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThr 59
Db GGCTCAGATGGGACTTCGCGGCCCAACAGCCCGCGCGCGCGCGCGCGAGATACG 17124
QY 60 GlnGlyAspGlyClnGlnSerAlaGlyGlyGlyProGlyAlaGluAlaLalleProAla 79
Db CAGGGCGCGCGGCAACAGAGTGGCGGAGCGCGCGCGCGGAGGAGCGATCCCGGCA 17184
QY 80 AlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99
Db GCAGCTGCTGCAGCGGTGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCG 17244
QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119
Db GCGCGCGCGGAGCGCGGAGGTGCACGAGCAAGCCATATACGCGCGCGCGCGCAAGCCC 17304
QY 120 ProTyrSerTyrIleAlaLeuAlaMetAlaLeuArgAspSerAlaGlyGlyArgLeu 139
Db CCTACTCGTACATCGCTCATCGCATCGGCATCCGCGACTCGCGCGCGCGCGCGCTTG 17364
QY 140 ThrLeuAlaGluLeuAsnGluTyrLeuMetGlyLysPhePhePhePheArgGlySerTyr 159
Db ACCTGCGGAGATCAACGAGTACCTCATGGCAAGTTCCTCTTTTCCGCGCGAGCTAC 17424
QY 160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValysVal 179
Db ACGGCTGCGCAACTCCGTGGCGCCCAACCTTCCTCAACGACTGCTTCGCAAGGTG 17484
QY 180 LeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSer 199
Db CTGCGGAGCCCTTCGCGCGCTGGGCGCAAGCACTACTGGATGCTCAACCCCAACAGC 17544
QY 200 GluTyrThrPheAlaAspGlyValPheArgArgArgArgArgLysLeuSerHisArgAla 219
Db GAGTACACCTTCGCCGACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17604
QY 220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaProPro 239
Db CCGGTCCCGCGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17664
QY 240 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259
Db CCGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17724
QY 260 AlaSerProAlaGlyLysPheSerSerPheAlaLeuAspSerIleLeuArgLysPro 279

US-10-087-080-32 (1-402) x AF225950 (1-4109)	
QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20	
DB 1737 ATGAAGTTGGAGGTGTCTGCTCCGCGGCGCCACGGGACAGCAGGAGTGCACCTG 1796	
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40	
DB 1797 GAGGCGCGCGGCGGAGCAGCAGCGCGCTCCCGCTGTCCGCGCGGAGAGACTCCCTG 1856	
QY 41 GlySerAspGlyAspCysAlaAla--LysProSerAlaGlyGlyGlyAlaArgAspThr 59	
DB 1857 GGCTCAGATGGGACTGCGCGGCGCCACAGCCCGCGCGCGCGCGCGCGCGGAGATCCG 1916	
QY 60 GlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAla 79	
DB 1917 CCGGCGCGCGGCGGAAACAGAGTGGGAGCGCGCGCGCGCGCGCGCGCGCGCGGCA 1976	
QY 80 AlaAlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99	
DB 1977 GCAGCTGCTGCAAGCGGTGTGGCGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCG 2036	
QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgProLysPro 119	
DB 2037 GCGCGCGCGGAGCGCGCGGAGGTGCACGACGACGACATATACGCGCGCGCGCGCG 2096	
QY 120 ProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 139	
DB 2097 CCTACTCGTACATCGCTCATCGCATGGCCATCGCGACTCGCGCGCGCGCGCGCTG 2156	
QY 140 ThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyr 159	
DB 2157 ACCTGCGCGGAGNTCAACAGTACCTCATGGCGAGTTCCTTTTCGCGCGCGCTAC 2216	
QY 160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysVal 179	
DB 2217 ACGGCTGCGCAACTCCGTGGCGCCACAACCTTCGCTCAAGACTGCTTCGTCAGGTG 2276	
QY 180 LeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSer 199	
DB 2277 CTGCGCGACCCCTCGCGGCGCTCGGCGCAGGACAACTACTGATGCTCAACCCCAACG 2336	
QY 200 GluTyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLysSerHisArgAla 219	
DB 2337 GAGTACACCTTCGCGACGGGTCTTCGCGCGCGCGCGCGCGCGCTCAGCCACCGCG 2396	
QY 220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro 239	
DB 2397 CCGGTCGCGCGCGCGGTGCGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2456	
QY 240 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259	
DB 2457 CCGCGCGCGCGCGCGCGCTCGCCCGCGATGCGCTCGCGCGCGCGCGCGCGCGCG 2516	
QY 260 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279	
DB 2517 GCCAGCGCGCGCGGCAAGTCTCCAGCTCTTCGCCATCGACAGCATCTCGCGCGCG 2576	
QY 280 PheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrPglyAlaAla 299	
DB 2577 TTCGAGCGCGCGCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2636	
QY 300 ProCysProProLeuProAlaPheProAlaLeuLeuLeuProAlaAlaProCysArgAlaLeu 319	
DB 2637 CCTGCGCGCGCGTGGCGCGGTTCGCGCGGTCTCTCCCGCGCGCGCGCGCGCGCG 2696	
QY 320 LeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 339	
DB 2697 CTGCGCGCTGTCGCGGTACGCGCGCGGAGCCCGCGCGCTGGCGCGCGCGCGCGCG 2756	
QY 340 ValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLys 359	
DB 2757 GTCCACCGACCGCGCGCGCTCTCTGCTTGACCTCTCCGCGCGCGCGCGCGCGCAAG 2816	
QY 360 ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAla 379	
DB 2817 CCCTCCGAGCA 2876	
QY 380 AlaLeuGlnAlaAlaLeuValArgAspProGlyProHisLeuSerTyrProValGluThr 399	
DB 2877 GCCTTCGAGCGCGCGCTCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2936	
QY 400 LeuLeuAla 402	
DB 2937 CTCTAGCC 2945	
RESULT 5	
AF154426	1513 bp DNA linear ROD 24-AUG-2001
LOCUS	Mus musculus HPH1 (Hfh1) gene, partial cds.
DEFINITION	
ACCESSION	AF154426
VERSION	AF154426.1 GI:8347144
KEYWORDS	
SOURCE	Mus musculus (house mouse)
ORGANISM	
REFERENCE	1 (bases 1 to 1513)
AUTHORS	Hong H.-K., Noveroske, J.K., Headon, D.J., Liu, T., Sy, M.S., Justice, M.J. and Chakravarti, A.
TITLE	The winged helix/forhead transcription factor Foxq1 regulates differentiation of hair in satin mice
JOURNAL	Genesis 29 (4), 163-171 (2001)
MEDLINE	21207067
PUBMED	11309849
REFERENCE	2 (bases 1 to 1513)
AUTHORS	Hong H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.
TITLE	Direct Submission
JOURNAL	Submitted (26-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
FEATURES	Location/Qualifiers
source	1. .1513
gene	/organism="Mus musculus"
mRNA	/mol_type="genomic DNA"
CDS	/strain="129"
	/db_xref="taxon:10090"
	<158. >1360
	/gene="Hfh1"
	<158. >1360
	/product="Hfh1"
	158. 1360
	/gene="Hfh1"
	/codon_start=1
	/product="Hfh1"
	/protein_id="AAF74524.1"
	/db_xref="GI:8347145"
	/translation="MKLEVFVPRAAHGDKMGSDLEAGSSDVPSPLSAGDLSLGSDDGCAANSAAAGAGDLEGGSSGGPSAODGPBTDSDRTQAAGACGAGVGGEGARSKPTTRRPKPPSYIALIAIRDSAGRLTAEINELMGKFPFRGSGVTGWRNSVRHNLSDNCFVKLRDPSRFGWKNYMLNPSEYTPADGVFRERKRLSHRTTVSAGSLRPEEPAGPGAGTTPQAPARSPARSFARQESSSPASKFSSSPALSDLSKPFRRSRDGSALGQVLPWGAAPCPFLRAYPALLPAPAGGALLPLCAYGASEPTLLASRGTEVQPAAPLALLAPLSTAPAKPFRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLSYVPEVTLA"
ORIGIN	
Alignment Scores:	
Pred. No.:	6,64e-36 Length: 1513
Score:	1597.00 Matches: 320
Percent Similarity:	82.27% Conservative: 14
Best Local Similarity:	78.82% Mismatches: 62
Query Match:	75.22% Indels: 10
DB:	10 Gaps: 6

US-10-087-080-32 (1-402) x AF154426 (1-1513)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 Db 158 ATGAATTTGGAGGTCTTCCTCCACGCGAGCCCAACGGGGACAAATGGCGAGGATCTG 217
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
 Db 218 GAGGGGGCGGCGAGCAGCGCTGCATCTCCACTGTCCGCGGCTGGTGACGACTCCTTA 277
 QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59
 Db 278 GGCTCAGACGGGAGCTGTGCACCAACAGCCCGCGCGGCGAGCGCGGGGATCTG 337
 QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro 78
 Db 338 GAAGGTGGCGCGCGAGGAGNATTGAGTGGCGCGCGCGAGCCCAAGACGGTCCGGAG 397
 QY 79 AlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGly 98
 Db 398 GCACTGATGACAGCAGA-----ACGCAAGGCTCCGCGGCGAGGCGCGTGC 442
 QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLys 118
 Db 443 CGGGCGCGTGGCGCGCGAGGGCGCGCGAGCCGCTACACGCGCGCGCCCAAG 502
 QY 119 ProProTyrSerTyrIleAlaAlaIleAlaMetAlaIleArgAspSerAlaGlyArg 138
 Db 503 CCCCATACTCTCATCGCTCTCATCGCCATGCGCATCCCGGACTCCGCGGGCGGACGC 562
 QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySer 158
 Db 563 CTGACACTGGCGAGATCAAGATACCTCATGCGCAAGTTCCCTTTTCCGGGGGAGC 622
 QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178
 Db 623 TACACGGGCTGGCGCAACTCCGTGGCCACCACTCTCGCTCAACGACTGTTTGGTCAAG 682
 QY 179 ValLeuArgAspProSerArgProTyrGlyLysAspSerTyrTyrMetLeuAsnProAsn 198
 Db 683 GTGCTGGCGGACCCCTCGCGCCCTGGGGAGGAGCAACTACTGGATGCTCAACCCCAAC 742
 QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLeuSerHisArg 218
 Db 743 AGCGAATACACTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGG 802
 QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237
 Db 803 ACCACAGTCTCCGGTCCGGGCTCGCGCGGAGGAAGCCCCACCGGACCTGCGGGACC 862
 QY 238 ProProProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGlu 257
 Db 863 CCGAGCCCGCGCGCGCGCGCTTCCTCCCGATCGCGGCTCGCGGCTCGCCAGAG 922
 QY 258 GluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArg 277
 Db 923 GAGCGCTCCAGCCCTCGCAGCAAGTTCTCCAGCTCTTCGCCATCGACGACTTCTCAGC 982
 QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGly 297
 Db 983 AAGCTTTTTCGAGCGCGCGCGAGCGGAGCTCGGCTCTGGGGGTGAGCTACCTCGGGC 1042
 QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317
 Db 1043 CGCGCTCCCTCGCGCGCTCGCGGCTATCCCGGCTCTTCCTTCGCGCGCGCGGTGGC 1102
 QY 318 AlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGlu 337
 Db 1103 GCTCTGCTACCGCTGTGCTTACCGCGCAAGCGAGGCTACGCTGCTGGCGGTGCGCGGG 1162
 QY 338 AlaGluValProThrAlaProLeuLeuLeuAlaProLeuProAlaAlaAlaAlaPro 357
 Db 1163 ACCGAGGTGACCGCGCGCG---CCCTTCTGTGCGCGCCCTCTCCACCGCGGCTCCA 1219

QY 358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376
 Db 1220 GCCACGCCATTCGAGTTCGGAGACCGCGCGCGCGCGACCTGACTGCCCCCTACGG 1279
 QY 377 LeuProAlaAlaLeuGlnAlaAlaLeuValaArgArgProGlyProHisLeuSerTyrPro 396
 Db 1280 CTGCCACGCGCTCGAGCGCGCGCGCTGGGTCCGGTCCGCGACCTGTCTCTACCG 1339
 QY 397 ValGluThrLeuLeuAla 402
 Db 1340 GTGGAGACTCTGCTAGCT 1357

RESULT 6

BC047155 2046 bp mRNA linear ROD 07-OCT-2003
 Mus musculus forkhead box Q1, mRNA (CDNA clone MGC:54661
 IMAGE:6489152), complete cds.

LOCUS

BC047155

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2046)

Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,

Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,

Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L.,

Schetter,T.E., Brownstein,M.J., Udwin,T.B., Toshikiyuki,S.,

Carrollin,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,

Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,

McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

Fahy,J., Helton,E., Kettner,M., Madan,A., Rodriguez,S.,

Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,

Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,

Schmerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 2046)

Strausberg,R.

Direct Submission

Submitted (21-FEB-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgi.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>Contact: nisc.mgc@nih.gov

Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,

Dierich,N.L., Granito,S., Guan,X., Gupta,J., Haghghi,P.,

Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Laric,P., Legaspi,R.,

Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,

McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 100 Row: n Column: 18
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6754191.

FEATURES

Location/Qualifiers
 1..2046
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="MGC:54661 IMAGE:6489152"
 /tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month old mouse, taken by biopsy."
 /clone_lib="NCI CGAP Mam2"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"

gene

1..2046
 /gene="Foxq1"
 /notes="synonym: FHF-1"
 /db_xref="LocusID:15220"
 /db_xref="MGI:1298228"
 481..1683
 /codon_start=1
 /product="forkhead box Q1"
 /protein_id="AAH47155.1"
 /db_xref="GI:28502982"
 /db_xref="LocusID:15220"
 /translations="MKLEVFVFAAHGDKVGLSGSSDVPSPLSAAGDLSLSDG
 DCAANSFAAGSGAGDLEGGSGSADQPEATDSDRTQASAGPCAGGVGGG
 EGARSFPTRRKPPYSYIALAMAIRDSAGGLTLAEINYLKGFPPFRGSYTGWR
 NSVHNLINDCFLVKLRDPSPWMDNMYMLNPSEYTFADGVFRKRKRLSHRTTV
 SAGSLPERAPPGAGTPOAPAPASPIARPSPARQERSSPASXFSFADSLISK
 PFRSRDGSALGVQVPMWGAAPCPPLRAYPALLPAAPGALLPLCAYGASEPTILASR
 GTVQPAAPLLAPLSTAPAPKPPFPETAGAAHLYCPLELTLQAAACCPGPHLS
 YPVETLLA"

CDS

misc_feature

817..1089
 /note="FH; Region: FORKHEAD"
 /db_xref="CDD:smart00339"

ORIGIN

Alignment Scores:

Pred. No.: 8.39e-36 Length: 2046
 Score: 1597.00 Matches: 320
 Percent Similarity: 82.27% Conservative: 14
 Best Local Similarity: 78.82% Mismatches: 62
 Query Match: 75.22% Indels: 10
 DB: 10 Gaps: 6

US-10-087-080-32 (1-402) x BC047155 (1-2046)

QY 1 MetLysLeuGluValPheValProAtrGAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 Db 481 ATGAAATTGGAGGTGTTTCGTCACCGCGCGACCCACCGGGGACAAATGGCAGCGATCTG 540
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaGlyAspAspSerLeu 40
 Db 541 GAGGGGCGCGACGACGACGATGTCCTCCATCTCCATGTCGCGGCTGGTACGACTCCCTTA 600
 QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaAtrAspThr 59
 Db 601 GGCTCAGCGGGGACTGTGCAGCCACACAGCCGCGCGGCGGACGCGCGCGGGATCTG 660
 QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyProGlyProGlyAlaGluAlaIlePro 78
 Db 661 GAAGGTGGCGCGGAGAGGAATTCGATGTGGCGGCGCGAGCCCAAGACGCTCCGGAG 720
 QY 79 AlaAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaGlyProGly 98
 Db 721 GCAACTGATGACACAGA-----ACGACGACCTCCGCGGCGAGGCGGTGC 765

RESULT 7

AL589738 187709 bp DNA linear ROD 29-JUN-2002
 LOCUS Mouse DNA sequence from clone RP23-322J11 on chromosome 13,
 DEFINITION complete sequence.
 ACCESSION AL589738
 VERSION AL589738.6 GI:15422196
 KEYWORDS HTG.

QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaAtrSerLysProThrThrArgAtrProLys 118
 Db 766 CGCGGCGGCGTGGCGGCGGCGAGGCGCGCGAGCAAGCCGTACACGCGCGGCCCAAG 825
 QY 119 ProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleAlaSerAlaGlyGlyArg 138
 Db 826 CCCCATCTCTCTACATCGCTCTCATCGCATCGCATCGCGACTCCGCGGCGGACGC 885
 QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySer 158
 Db 886 CTGACACTGGCCGAGATCAACGAGTACTCTATGGGCAAGTTCCTTTTCCGGGCGAGC 945
 QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178
 Db 946 TACACGGGCTGGCGCAACTCCGTGGCCCAACAACCTCTCGCTCAACGACTGTTTCGTCAAG 1005
 QY 179 ValLeuArgAspProSerArgProTyrGlyLysAsnAsnTyrTrpMetLeuAsnProAsn 198
 Db 1006 GTGTGCGGACCCCTCGCGCCCTGGGCGAGGACACTACTGGATGCTCAACCCCAAC 1065
 QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArg 218
 Db 1066 AGCGAATACACTTCGCGGACGCGGCTCTTCGCGCGCGCAAGCGCTCAGCACCCG 1125
 QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237
 Db 1126 ACCACAGTCTCCGCTCGGCTCGCGCGGAGAGAGCCACCCGCGACTCGCGGAGCC 1185
 QY 238 ProProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlu 257
 Db 1186 CCGCAGCCCGCGCGCGCGCTCTCCCGACTCGCGGCTCGCGGCTCGCAGGAG 1245
 QY 258 GluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArg 277
 Db 1246 GAGCGCTCCAGCCCTCGCGCAAGTCTTCAGCTCTTCGCGCATCGACAGCATCTCAGC 1305
 QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly 297
 Db 1306 AAGCTTTTCGACGCGCGCGCGGACTCGGCTCTGGGGGTGACAGTACCCCTGGGGC 1365
 QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317
 Db 1366 CCGCTCCCTCGCGCGCGCTTCGCGGCTATCCCGGCTCTTCGCGGCGCGCGGTGGC 1425
 QY 318 AlaLeuLeuProLeuCysAlaTyrGlyValGlyGluProAlaArgLeuGlyAlaArgGlu 337
 Db 1426 GCTCTGCTACCGCTCTGTGCTTACGGCGCAAGCGAGCTACGCTGCTGGGTGCGCGGG 1485
 QY 338 AlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaPro 357
 Db 1486 ACCGAGGTGACGCGCGCGCG---CCCTTCTGTGGCGGCGCTCTCCACGCGGCTCCA 1542
 QY 358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376
 Db 1543 GCCAAGCATTCCTCGAGGTCCCGAGACCGCGCGCGCGGCGGCGGCTGCTGCTGCTACCG 1602
 QY 377 LeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrPro 396
 Db 1603 CTGCCACCGGCTTCGAGCGGCGAGCGGCTCGGCTCGCGTCCGCGTCCGCTGCTTACCCG 1662
 QY 397 ValGluThrLeuLeuAla 402
 Db 1663 GTGGAGACTCTGCTAGCT 1680

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 187709)

AUTHORS Phillimore,B.

TITLE Direct Submission

JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquy@sanger.ac.uk

COMMENT On Sep 3, 2001 this sequence version replaced gi:13445573. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-322J11 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: <http://mrcseq.har.mrc.ac.uk>
Contact: mouse@har.mrc.ac.uk

Location/Qualifiers
1..187709
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/clone="RP23-322J11"
/clone_lib="RPCI-23"

ORIGIN

Alignment Scores: 2.76e-34 Length: 187709
Pred. No.: 1597.00 Matches: 320
Score: 82.27% Conservat: 14
Percent Similarity: 78.82% Mismatches: 62
Best Local Similarity: 75.22% Indels: 10
Query Match: 10 Gaps: 6
DB:

US-10-087-080-32 (1-402) x AL589738 (1-187709)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20

DB 135706 ATGAATTTGGAGGTTCCTCCACGCGCAGCCACCGGGGACAAAATGGCGCGATCTG 135765

QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40

DB 135766 GAGGGGGCGCGCAGCAGCGAGCTGTCATCTCCATGTCCCGCGCTGGTGAGACTCTTTA 135825

QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59

DB 135826 GGCTCAGCGGGAGTGTGCGAGCCACAGCCCGCGCGCGCGCGCGCGGATCTG 135885

QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluAlaIlePro 78

DB 135885

RESULT 8
AF010405
LOCUS

AF010405 4763 bp DNA linear ROD 20-JAN-2000

135886 GAAGTGGCGCGCGCAGAGGAATTCGAGTGGCGGGCCGAGCGCCAAAGACGGTCCGGAG 135945

QY 79 AlaAlaAlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGly 98

DB 135946 GCAACTGATGACAGCAGA-----ACGAGGCGCTCCGCGCAGCGCGCGTGC 135990

QY 99 AlaGlyGlyAlaGlySerGlyGlyAlaArgSerLysProTyThrArgArgProLys 118

DB 135991 GCGGGCGCGTGGCG 136050

QY 119 ProProTyThrSerTyIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArg 138

DB 136051 CCCCATCTCTCTACTCGCTCTCATCGCATCGGCATCCCGGATCTCCGCGCGCGAGCG 136110

QY 139 LeuThrLeuAlaGluIleAsnGluTyLeuMetGlyLysPheProPhePheArgGlySer 158

DB 136111 CTGACACTGGCGCGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTCCGGGCGAGC 136170

QY 159 TyThrGlyTyThrArgAspSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178

DB 136171 TACACGGGCTGGCGCAACTCCCGTGGCGCCACAACTCTCGTCAACGACTGTTTCGTCAAG 136230

QY 179 ValLeuArgAspProSerArgProTyThrGlyLysAspAsnTyThrMetLeuAsnProAsn 198

DB 136231 GTGCTGGCGAGCCCTTCGCGGCCCTGGGGCAGGAGCACTACTGATGCTCAACCCCAAC 136290

QY 199 SerGluTyThrPheAlaAspGlyValPheArgArgArgArgGlyArgLeuSerHisArg 218

DB 136291 AGCAATATACCTTCGCGCGCGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136350

QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237

DB 136351 ACCACAGTTCGCGGTCCGCGGTGGCGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 136410

QY 238 ProProProAlaProAlaProAlaProAlaProArgMetArgSerProAlaArgGlnGlu 257

DB 136411 CCGAGCG 136470

QY 258 GluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArg 277

DB 136471 GAGCGCTCCAGCCCTCGCAGCAAGTTCCTCCAGTCTTCGCGCATCGACAGCAATCTCAGC 136530

QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly 297

DB 136531 AGCGCTTTTCGAGCG 136590

QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317

DB 136591 GCGCTCCCTGCG 136650

QY 318 AlaLeuLeuProLeuCysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGlu 337

DB 136651 GCTCTGCTACCGCTCTGTGCTTACGGCGCAGGAGCGCTACGCTGTGCGCGCGCGCG 136710

QY 338 AlaGluValProProThrAlaProProLeuLeuLeuAlaProleuProAlaAlaPro 357

DB 136711 ACCGAGGTGCG 136767

QY 358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyCysProLeuArg 376

DB 136768 GCCAACCATTCGAGGTTCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136827

QY 377 LeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyPro 396

DB 136828 CTGCCACGCGCTCGAGCG 136887

QY 397 ValGluThrLeuLeuAla 402

DB 136888 GTGGAGACTGTGCTAGCT 136905

DEFINITION Mus musculus fork head transcription factor (Hfh-1L) gene, complete.

ACCESSION AF010405

VERSION AF010405.2 GI:6716869

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 4763)

AUTHORS Frank, S. and Zoll, B.

TITLE Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal location, and expression in adult and embryonic kidney

JOURNAL DNA Cell Biol. 17 (8), 679-688 (1998)

MEDLINE 98392851

PUBMED 9726250

REFERENCE 2 (bases 1 to 4763)

AUTHORS Frank, S. and Zoll, B.

TITLE Direct Submission

JOURNAL Submitted (19-JUN-1997) University of Goettingen, Institute of Human Genetics, Gosselerstr 12d, Goettingen 37073, Germany

REFERENCE 3 (bases 1 to 4763)

AUTHORS Pasche, B., Bieller, A. and Zoll, B.

TITLE Direct Submission

JOURNAL Submitted (19-JAN-2000) University of Goettingen, Institute of Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany

REMARK Sequence update by submitter

COMMENT On Jan 20, 2000 this sequence version replaced gi:2997586.

FEATURES

source

1..4763

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="129"

/db_xref="taxon:10090"

/chromosome="13"

/tissue_type="kidney"

<2467..33669

/gene="Hfh-1L"

/notes="HNF-3/fork head homologue-1 like"

<2467..33669

/gene="Hfh-1L"

/product="fork head transcription factor"

2467..3669

/gene="Hfh-1L"

/notes="member of HNF-3/fork head transcription factor family"

/codon_start=1

/product="fork head transcription factor"

/protein_id="AAC12973.2"

/db_xref="GI:6716870"

/translation="MKLEVPRAAHGDMGSDLEGAGSDVPSPLSRAAGDSDLSGDG DCANSPAGSGAGDLEGGERNSGSGSAQDGPATDDSDRTQASAPGAGGVGG EGARKEPYTRPRKPYSTIALIAIRDSAGRLTAEINEYLMGKFPFRSGYTGWR NSVHNLSLNDCKVKLDRSPRWKNDYMLNPNSEYTFADGVFRFRKRLSHRTV SASGLRDEEPAGPTPOAAPSPARSPIARSPORBERSSPASKFSSFAIDSLSK PFSRRRDGSDALGVQFWGAAPCPPLRAYPALLPAPGALLPLCAYGASEPTLLASR GTVQPAAPALLPLSTAAKPFPRGPETAGAAHLYCLRLPLTALQAAAACGPGPHLS YPVELLLA"

2782..3123

/gene="Hfh-1L"

/note="encodes DNA binding domain"

ORIGIN

Alignment Scores:

Prod. No.: 2,66e-35 Length: 4763

Score: 1589.00 Matches: 319

Percent Similarity: 82.02% Conservative: 14

Best Local Similarity: 78.57% Mismatches: 63

Query Match: 74.85% Indels: 10

DB: 10 Gaps: 6

US-10-087-080-32 (1-402) x AF010405 (1-4763)

QY 1 MetLysIeuGluValPheValProArgAlaAlaHisGlyAspLysGingLysSerAspLeu 20

Db 2467 ATGAATTTGAGGTGTTTCCTCCACGCGCAGCCACCGGGACAAATAATCGGACGCACTG 2526

QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40

Db 2527 GAGGGGCGCGCAGCAGCAGCGTCCATCTCCATGTCCGCGGTGTGTGACGACTCCTTA 2586

QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59

Db 2587 GGTCTACAGCGGGGACTGTGCGAGCAACAGCCCGCGCGGCGGCGGCGGCGGATCTG 2646

QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluAlaIlePro 78

Db 2647 GAAGTGTGGCGCGCAGAGAGAAATTGAGTGGCGGCGGAGCGCCCAAGACGCTCCGAG 2706

QY 79 AlaAlaAlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGly 98

Db 2707 GCAACTGATGACAGCAGA-----ACGCGAGCCCTCCGCGGCGGCGGCGGTGC 2751

QY 99 AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgProLys 118

Db 2752 CGCGCGCGGTGGCG 2811

QY 119 ProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArg 138

Db 2812 CCCCATCTCTCTATCGCTCTCATCGCATGCGCATCGCGACTCGCGGCGGCGGACGC 2871

QY 139 LeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySer 158

Db 2872 CTGACACTGCGCGAGATCAACGAGTACTCATGGCGAAGTCCCTTTTCCGCGGCG 2931

QY 159 TyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLys 178

Db 2932 TACACGGGCTGGCGCAATCTCGTGGCGCAACACCTCTCGTCAACGACTGTTGCTCA 2991

QY 179 ValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsn 198

Db 2992 GTGCTGCGGACCCCTCGCGCCCTGGGCGCAAGCAACTACTGATGCTCAACCCCAAC 3051

QY 199 SerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArg 218

Db 3052 AGCGAATACACTTTCGCGCAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTCAGCACCG 3111

QY 219 AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---Ala 237

Db 3112 ACCACAGTCTCCGGTCCGGCTCGCGCGGAGAGACCCACCCGCGACTGCGCGGAC 3171

QY 238 ProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGlu 257

Db 3172 CCGCAGCGCGCGCGCGCGCGCTCTCCCGATCGCGCGCTCGCGGCTCGCGCAGAG 3231

QY 258 GluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArg 277

Db 3232 GAGCGCTCCAGCCCTCGCGAGCAATTCTCCAGCTCTTCGCGCATCGACGACTTCTCAG 3291

QY 278 LysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGly 297

Db 3292 AAGCTTTTCGCGCGCGCGCGCGCGCTCTCGGGGTGCGAGTACCTCGCGGCG 3351

QY 298 AlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArg 317

Db 3352 GCGCTCCCTCGCGCGCGCTTCGCGCGCTATCCCGCGCTCTTCGCGGCGCGCGGTCG 3411

QY 318 AlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGlu 337

Db 3412 GCTCTGCTACCGCTCTGTGCTTACGGCGCAGCGAGCTACGCTGCTGCGTCTCGCGGG 3471

QY 338 AlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaPro 357

Db 3472 ACCGAGGTGCACCGCGCGCG---CCCTTCTGTGCGCGCGCTCTCCACCGCGGCTCCA 3528

QY 358 AlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeuArg 376

```

Db      3529  GCGAAGCCATTCGAGGTCCGAGACCGCGCGCGCGCGACCTGTACTGCCCCCTACGG 3588
Qy      377  LeuProAlaAlaLeuGluAlaAlaLeuValargProGlyProHisLeuSerTy-Pro 396
Db      3589  CTGCCACGCGCCTCGACGCGCAGCGCGCTCGCGTCCCGTCCGACCTCTCTCTACCCG 3648
Qy      397  ValGluThrLeuLeuAla 402
Db      3649  GTGGAGACTCTGCTAGCT 3666

RESULT 9
AF153193
LOCUS      1576 bp      DNA      linear      ROD 24-AUG-2001
DEFINITION Rattus rattus winged helix/forkhead transcription factor HFH1
            (Hfh1) gene, complete cds.
ACCESSION AF153193
VERSION    AF153193.1 GI:8132282
KEYWORDS   Rattus rattus (black rat)
SOURCE     Rattus rattus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 (bases 1 to 1576)
AUTHORS   Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
            Justice,M.J. and Chakravarti,A.
TITLE     The winged helix/forkhead transcription factor Foxq1 regulates
            differentiation of hair in satin mice
JOURNAL   Genesis 29 (4), 163-171 (2001)
MEDLINE   21207067
PUBMED    11309849
REFERENCE 2 (bases 1 to 1576)
AUTHORS   Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
TITLE     Direct Submission
JOURNAL   Submitted (20-MAY-1999) Genetics, Case Western Reserve University,
            10900 Euclid Ave., Cleveland, OH 44106-4955
FEATURES
    source
        1..1576
            /organism="Rattus rattus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10117"
            <95..>1297
            /gene="Hfh1"
            <95..>1297
            /gene="Hfh1"
            /product="winged helix/forkhead transcription factor HFH1"
            95..1297
            /codon_start=1
            /protein_id="AAF73234.1"
            /db_xref="GI:8132283"
            /translation="MKLRFAPRAAHGDKMSDLEGAGSSDVPSPLSAAGDSDLGSDG
            DCAANSAPAGAVDLGGGERSNGSGASTQDDFDVTDGRTQASVPVPGCVGGG
            EGAKPRTFRPKFPYSYIALTAIRDSAGGRUTLAEINEYLMGKPPFFRGSYGWR
            NSVRNLGNCVFLVDRPDKWKNYMLNPNSEYTFADGVRFRRLKLSHRTTV
            SASGLRPEAPGPAGTQPAPATGSSPIARQEGSSPASKFSFSAIDSLSK
            PPRSRDGPALGPVLPWAAAPCPPLRAYPALLPASGGALLPLCAYGAGEPTLASR
            GAEPVAPALPLAIPSTAAPKPRGPTAGAHLYCPLRLPTALQAACGPGPHLS
            YRVETLLA"
    gene
    mRNA
    CDS

```

ORIGIN

Alignment Scores:

```

Pred. No.:      3,5e-34      Length:      1576
Score:           1534.00      Matches:      311
Percent Similarity: 79.90%      Conservative: 15
Best Local Similarity: 76.23%      Mismatches: 68
Query Match:      72.26%      Indels:      14
DB:               10          Gaps:         7

```

US-10-087-080-32 (1-402) x AF153193 (1-1576)

```

Qy      1  MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db      95  ATGAATTTGGAGTATTTGCCCCACGCGCAGCCACCGGGACAAGATGGCAGTGCACCTG 154
Qy      21  GluGlyAlaGlyClySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
Db      155  GAGGGGGCGCGGACGACGACGTCCTCTCGCGCTCTCGCGGCTGGCAGCAGCTCCTTA 214
Qy      41  GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyAlaArgAspThr 59
Db      215  GGCTCTGACGGGACTGTGCAGCCACAGCCCGCGCGCGGCGAGAGCGCGGTGATCTG 274
Qy      60  Gln---GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaLeuPro 78
Db      275  GAAGGCGCGCGCGCGAGAGGAATTCGAGTGGCGGGCGAGCACCACAGAC----- 325
Qy      79  AlaAlaAlaAlaAlaValValAlaGluGlyAla-----GluAlaGlyAlaAlaGly 96
Db      326  -----GATCCCGAGGTGACCGATGGCAGCAGAACGACGAGGCTCCCGGTGGG 373
Qy      97  ProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArg 116
Db      374  CCGTGGCGGCGGACGCGTGGCGCGGTGAGGGCGCGCGCAGCAAGCGGTACACGCGCGG 433
Qy      117  ProLysProProTyrSerTyrTrileAlaLeuAlaMetAlaAlaIleArgAspSerAlaGly 136
Db      434  CCCAAGCCCCCTACTTCTTACATCGCACTATTCGCATTCGCCATTCGCCGACTCCGCGG 493
Qy      137  GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArg 156
Db      494  GGACGCGCTCGCGCGACGCTCAACGAGTACTCATGGGCAAGTTCCTCTTTTCCCG 553
Qy      157  GlySerTyrThrClyTyrPargHsnSerValArgHisAsnLeuSerLeuAsnGlyPhe 176
Db      554  GGAGCTTACAGCGCGTGGCGCAACTCCGTGGCGCAACCTCTCGCTCAACGACTGTTTC 613
Qy      177  ValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsn 196
Db      614  GTCAAGGTGCTCGCGCACCCCTCGCGGCTGGGGCAAGGACAATTAATGATGCTCAAC 673
Qy      197  ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSer 216
Db      674  CCCAAGCAGCAATACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCAGCGCCTCAGC 733
Qy      217  HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
Db      734  CACCGACACACAGTCTCCGATCGGGGCTACGGCGGAGGAGGCCACCCGAGCCTGCG 793
Qy      237  ---AlaProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArg 255
Db      794  GGGACCCCGCAGCGCGCGCCACCGCGGCTCTCTCCCAATCGCGGCGCTCGCCCGCTCG 853
Qy      256  GlnGluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIle 275
Db      854  CAGAGAGGGGTCCAGCCCGCGGAGCAAGTTCAGTCTCTTCCGCTCCGACGACATC 913
Qy      276  LeuArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGln 295
Db      914  CTCAGCAAGCCGTTTCGACGCGCGCGCAGCGCGACCCGCGCTCTGGGGGTGTCAGCTACC 973
Qy      296  TrpGlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaPro 315
Db      974  TGGAGCGCTGCTCCCTGCGCGCGCTATCCCGCGCTCTCTCCGCGCTGCTCC 1033
Qy      316  CysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla 335
Db      1034  GCGGCTGCCCTGCTGCGCTGTGTGCTTACGGCGCGGGGAGGCCACGCTGCTGGGCTCG 1093
Qy      336  ArgGluAlaGluValProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAla 355
Db      1094  CGCGGGCGGAGGTGAGCCCGCGGCG---CCCTCTTGTGCGCGCCCTCTCCACCGCG 1150
Qy      356  AlaProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysPro 374

```

```

1151 GCGCCAGCAAGCATTTCAGGTCCGAGACCGCGCGCGCGCACCTGTACTGCCCC 1210
|||||
375 LeuAqGLeuProAlaLalaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSer 394
|||||
1211 CTACGGCTGCCACCGCCCTGAGGCGCGCGCGCGCTCGGTCGCGGTCCGACCTGTCC 1270
|||||
395 TyrProValGluThrLeuLeuAla 402
|||||
1271 TACCGGGTGGAGACGCTGTAGCT 1294
|||||

RESULT 10
AC119497
LOCUS
DEFINITION
Rattus norvegicus clone CH230-521F8, *** SEQUENCING IN PROGRESS
***, 4 unordered pieces.
AC119497
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 195184)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,B., Benahmed,F.,
Biswal,K., Blair,J., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geier,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haviak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulvik,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebowitz,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwankweliemeh,O., Okwuonu,G., Olarunpungoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Pioppoer,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojes,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,U., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

```

```

Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 195184)
Worley,K.C.
Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195184)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23195197.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWQI
Center clone name: CH230-521F8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 158913 bases at least Q40
Consensus quality: 162109 bases at least Q30
Consensus quality: 164133 bases at least Q20
Estimated insert size: 161675; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16506: contig of 16506 bp in length
* 16507 16606: gap of unknown length
* 16607 180638: contig of 164032 bp in length
* 180639 180738: gap of unknown length
* 180739 182447: contig of 1709 bp in length
* 182448 182548: gap of unknown length
* 182548 195184: contig of 12637 bp in length.
* Location/Qualifiers
* 1..195184
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clones="CH230-521F8"
* 466..1181
* /note="clone boundary
* clone end:T7
* site:
* end sequence:BZ266854"
* 1768..2639
* misc_feature
* misc_feature

```



```

/note="clone boundary
clone_end:Sp6
site:
end_sequence:BZ267695"
misc_feature
16607..17640
/note="wgs end extension
clone_end:Sp6"
misc_feature
71528..72576
/note="wgs end extension
clone_end:Sp6"
misc_feature
76693..78955
/note="wgs end extension
clone_end:Sp6"
misc_feature
86077..87329
/note="wgs end extension
clone_end:Sp6"
misc_feature
106805..108354
/note="wgs end extension
clone_end:Sp6"
misc_feature
129622..130958
/note="wgs end extension
clone_end:Sp6"

ORIGIN
Alignment Scores:
Pred. No.: 1.45e-32 Length: 195184
Score: 1534.00 Matches: 311
Percent Similarity: 79.90% Conservative: 15
Best Local Similarity: 76.22% Mismatches: 68
Query Match: 72.22% Indels: 14
DB: 2 Gaps: 7

US-10-087-080-32 (1-402) x AC119497 (1-195184)
QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db 136153 ATGAATTTGGAGTATTTCCTCCACCGCAGCCACGGGACAGATGGGAGTGACTG 136212
QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
Db 136213 GAGGGGGCGGCGAGCAGCGACGTGCATCTCGCTGTCTCGGGGTGGGAGACTCTCTTA 136272
QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyGlyValaArgAspThr 59
Db 136273 GGCTCTGACGGGGACTGTGCACCAACAGCCCGCGGGCGGCGAGCGCCGTGATCTG 136332
QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaLeuPro 78
Db 136333 GAAGCGGGCGGCGGAGAGGATTCGAGTGGCGGGCGGAGCACCACAGAC----- 136383
QY 79 AlaAlaAlaAlaAlaValValaGluAlaGlyAla-----GluAlaGlyAlaAlaGly 96
Db 136384 -----GATCCCGAGGTGACCGATGTCGACGACGACGACGACGCGCTCCCGGTGGG 136431
QY 97 ProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProGlyThrArgArg 116
Db 136432 CCGTGGCGGGGAGCGGTGGCGGGGTGAGCGGGCGGCGGCGGACGACGCGGTACAGCGGGG 136491
QY 117 ProLysProProGlySerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 136
Db 136492 CCCAAGCCCCCTACTCTTACATCGACATCATCGGCATCGCCATCCGCGACTCCGCGGGC 136551
QY 137 GlyArgLeuThrIleuAlaGluIleGluTyrLeuMetGlyLysPheProPhePheArg 156
Db 136552 GGACGGCTGACGCTGGCCGAGATCAACGAGTACCTCATCGGGCAAGTTCCTCTTTTCGGG 136611
QY 157 GlySerTyrThrGlyTrpArgAsnSerValaArgHisAsnLeuSerIleuAsnAspCysPhe 176
Db 136612 GGCAGCTACACGGCTGGCGCAACTCCGTGCGCCACACACTCTCGCTCAACGACTGTTTC 136671
QY 177 ValLysValLeuArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsn 196
Db 136672 GTCAAGGTGCTGGCGGACCCCTCGCGGCGCTGGGGCAAGGACAAATTACTGGATGCTCAAC 136731

```

197 ProAsnSerGluTyrThrPheAlaAspGlyValPheAlaArgArgArgArgArgArgArg 216

136732 CCCAACAGCAATACACCTTCGCCACCGGGGTCTTCGCGCGCGCGCAAGCGCTCAGC 136791

217 HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236

136792 CACCGGACCACAGTCTCCGCATCGGGGTACGCGCGGAGAGAGCCACCCGAGCTGG 136851

237 ---AlaProProAlaProAlaProAlaProAlaProAlaProAlaProAlaProAla 255

136852 GGGACCCCGCAGCCCGGCCACCCCGCGGCTCTCCCAATCGCGGCTCGCCGCTCGC 136911

256 GlnGluGluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIle 275

136912 CAGGAGGAGGCTCCAGCCCGCGGAGCAAGTCTCCAGCTCTCCGCATCGACAGCATC 136971

276 LeuArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGln 295

136972 CTCAGCAAGCCGTTTCGACGCGCGCGGACCGCGGCTCTGGGGGTGAGCTACCC 137031

296 TrpGlyAlaAlaProCysProLeuProAlaPheProAlaLeuLeuProAlaAlaPro 315

137032 TGGAGCGCTGCTCCCTGCGCGGCTGCGCGCTATCCCGGCTCTTCCCGCTCGTCC 137091

316 CysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla 335

137092 GCGCGTGCCTGCTGCTGCTGTACGCGCGCGGAGCCGCGGCTGCTGGGCTGCG 137151

336 ArgGluAlaGluValProThrAlaProProLeuLeuLeuAlaProLeuProAlaAla 355

137152 CGCGGGCGGAGGTGAGCGCGCGG---CCCTGTGCTGGCGCCCTCTCCACCGG 137208

356 AlaProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysPro 374

137209 GCCCCAGCAAGCAATTCGAGTTCGAGACCGCGCGCGGCGGCTACTGCTGCTGCT 137268

375 LeuArgLeuProAlaAlaLeuGlnAlaLeuValArgArgProGlyProHisLeuSer 394

137269 CTACGGTGGCCACGCGCTTGGAGGGCGCGCGGCTGCGGCTCGGCTCGGCTGCTG 137328

395 TyrProValGluThrLeuLeuAla 402

137329 TACCGGTGGAGACGCTGCTAGCT 137352

RESULT 11

RATHFH1 1760 bp mRNA linear ROD 22-AUG-1995

LOCUS Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete cds.

DEFINITION L13201 GI:951067

VERSION L13201.1

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1760)

AUTHORS Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lai,E. and Costa,R.H.

TITLE Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)

MEDLINE 93248207

PUBMED 7683413

COMMENT On Aug 22, 1995 this sequence version replaced gi:550512.

FEATURES

source

Location/Qualifiers

1..1760

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

ORIGIN

Pred. No.:	7.5e-31	Length:	1760
Score:	1412.50	Matches:	300
Percent Similarity:	77.64%	Conservative:	16
Best Local Similarity:	73.71%	Mismatches:	78
Query Match:	66.53%	Indels:	17
DB:	1	Gaps:	6

1	Met	Ly	Ser	Leu	Glu	Val	Phe	Val	Pro	Arg	Ala	Ala	Ile	scd	Ly	Asp	Ly	Ser	Asp	Leu	20	
207	AT	GAA	TTGG	AGTG	-TT	CGT	CCC	ACG	CGC	GAG	-CAC	GGG	GAC	AAA	ATG	GGC	GAC	GCA	CTCTG	264		
21	Gl	u	Gly	I	Ala	Gly	Ser	Asp	Ala	Pro	Ser	Pro	Leu	Ser	Ala	Ala	Gly	Asp	Asp	Ser	Leu	40
265	GAG	GGG	CGG	CGG	CAG	CAG	CGC	AGC	AGT	GCC	ATC	TCC	ACT	GTC	CGG	GCT	GGT	GAC	GACT	CTCTTA	324	
41	Gly	Ser	Asp	Gly	Asp	Cys	Ala	Ala	---	Ly	P	Ser	Ser	Ala	Gly	Gly	Gly	Ala	Arg	Asp	Thr	59
325	G	GCT	CAG	ACG	GGA	CTG	TGC	AGC	CAAC	AGC	CCG	CGC	CGC	GAG	CGC	GCG	CGG	GGA	CTCTG	384		
60	Gln	---	Gly	Asp	Gly	Glu	Gln	Ser	Ala	Gly	Gly	Gly	Pro	Gly	Ala	Glu	Glu	Ala	Ile	Pro	78	
385	GA	AGT	TGG	CGG	CGG	CAG	GAG	GA	ATTC	GAT	GCG	GCG	GCG	GAG	CAC	CCCAAGAC	-----	435				
79	Ala	Ala	Ala	Ala	Ala	Val	Val	Ala	Glu	Gly	Ala	---	---	Glu	Ala	Gly	Ala	Ala	Gly	96		
436	-----	---	---	---	---	GAT	CCG	AGG	TAC	CGG	ATC	GAT	GGC	AGAG	ACG	GAG	GCT	CCCCG	GTGGG	483		
97	Pro	Gly	Ala	Gly	Gly	Ala	Gly	Ser	Gly	Glu	Gly	Ala	Arg	Ser	Ly	Ser	Pro	Thr	Arg	Arg	116	
484	CG	TG	CGG	CGG	CAG	CGT	GCG	GCG	GCG	GCG	GAG	CGC	GCG	GAG	CAAG	CGG	TAC	ACG	CGG	CGG	543	
117	Pro	Ly	Ser	Pro	Pro	Tyr	Ser	Tyr	Ile	Ala	Leu	Ile	Ala	Met	Ala	Arg	Asp	Ser	Ala	Gly	136	
544	CC	CA	AGC	CC	CC	CT	CT	CT	CT	AT	CGC	ACT	AT	CGC	AT	CGC	ACT	CCG	CGG	CGG	603	
137	Gly	Arg	Leu	Thr	Leu	Ala	Gly	Ile	Asn	Gly	Tyr	Leu	Met	Gly	Ly	Ser	Phe	Pro	Phe	Arg	156	
604	GG	AC	CC	CT	TGA	CGCT	TGG	CGG	AGT	CAAC	GAG	TAC	CTC	TAT	GGC	CAAGT	T	CCCT	TTT	CCCG	663	
157	Gly	Ser	Tyr	Thr	Gly	Trp	Arg	Asn	Ser	Val	Arg	Ile	Asn	Leu	Ser	Leu	Asn	Asp	Cys	Phe	176	
664	G	G	C	G	C	T	A	C	A	G	G	G	C	A	A	C	C	T	T	C	G	723
177	Val	Ly	Ser	Val	Leu	Arg	Asp	Pro	Ser	Arg	Pro	Trp	Gly	Ly	Ser	Asn	Tyr	Trp	Met	Leu	Asn	196
724	GT	CA	AGG	TC	TCT	CG	CGA	AC	CC	CT	CG	CGC	CC	CT	TGG	GCG	AAG	CAC	CAAT	TACT	GAT	783

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jun 4, 2000 this sequence version replaced gi:7655991.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10103

Center clone name: 403_I_20

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 173395 bases at least Q40

Consensus quality: 177417 bases at least Q30

Consensus quality: 179105 bases at least Q20

Insert size: 179300; agarose-fp

Insert size: 180452; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2222: contig of 2222 bp in length

* 2223 2222: gap of 100 bp

* 2323 3785: contig of 1463 bp in length

*

3786 3885: gap of 100 bp
 3886 6753: contig of 2868 bp in length
 6754 6853: gap of 100 bp
 6854 10036: contig of 3183 bp in length
 10037 10136: gap of 100 bp
 10137 14537: contig of 4401 bp in length
 14538 14637: gap of 100 bp
 14638 21177: contig of 6540 bp in length
 21178 21278: gap of 100 bp
 21279 30240: contig of 8963 bp in length
 30241 40123: gap of 100 bp
 40124 40223: contig of 9783 bp in length
 40224 47939: contig of 7776 bp in length
 47940 48039: gap of 100 bp
 48040 58235: contig of 10136 bp in length
 58236 58335: gap of 100 bp
 58336 67041: contig of 8706 bp in length
 67042 67141: gap of 100 bp
 67142 75921: contig of 8780 bp in length
 75922 76021: gap of 100 bp
 76022 86527: contig of 10506 bp in length
 86528 86627: gap of 100 bp
 86628 101334: contig of 14707 bp in length
 101335 101434: gap of 100 bp
 101435 117630: contig of 16196 bp in length
 117631 117730: gap of 100 bp
 117731 138293: contig of 20563 bp in length
 138294 138393: gap of 100 bp
 138394 157841: contig of 19448 bp in length
 157842 157942: gap of 100 bp
 157943 182152: contig of 24211 bp in length.

FEATURES

source

1. 182152
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosomes="6"
 /map="6"
 /clone="RP11-403I20"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature 1. 2222

misc_feature 2323..3785

misc_feature 3886..6753

misc_feature 6854..10036

misc_feature 10137..14537

misc_feature 14538..21177

misc_feature 21278..30240

misc_feature 30341..40123

misc_feature 40224..47939

misc_feature 48100..58235

misc_feature 58336..67041

misc_feature 67142..75921

misc_feature 76022..86527

misc_feature 86628..101334

misc_feature 101435..117630

misc_feature 117631..157942

misc_feature 157943..182152

misc_feature 182153..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

misc_feature 24212..24211

```

misc_feature 117731..138293
              /note="assembly_fragment"
misc_feature 138194..157841
              /note="assembly_fragment"
misc_feature 157942..182152
              /note="assembly_fragment
              clone_end.T7
              vector_side:right"

ORIGIN
Alignment Scores:
Pred. No.: 3.39e-28 Length: 182152
Score: 1372.00 Matches: 255
Percent Similarity: 99.22% Conservative: 2
Best Local Similarity: 99.22% Mismatches: 2
Query Match: 64.63% Indels: 0
DB: 2 Gaps: 0

US-10-087-080-32 (1-402) x AC067929 (1-182152)

Qy 146 GluTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTPAArgAnSer 165
Db 10036 GAGTACTCATGGCGAAGTTCCTTTTCGGGGGAGCTACACGGCTGGCGCACTCC 9977

Qy 166 ValArgHisLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArg 185
Db 9976 GTGGCCACACCTTCGCTCAGCACTGCTTCGTCAGGTGCTGCGGACCCCTCGCGG 9917

Qy 186 ProTyrGlyLysAspAnTyrTyrMetLeuAsnProAnSerGluTyrThrPheAlaAsp 205
Db 9916 CCTTGGGCGAGGACAACTACTGATGCTCAACCCCAAGCAGTACACTTCGGCGAC 9957

Qy 206 GlyValPheArgArgAArgLysArgLeuSerHisArgAlaProValProAlaProGly 225
Db 9856 GGGGTCTTCGCCCGCGCGAGCGCTCAGCCACCGCGCGCGTCCCGCGCGCGCGG 9797

Qy 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaPro 245
Db 9796 CTGGCGCGGAGGAGGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 9737

Qy 246 AlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLys 265
Db 9736 GCCTCGCGCGCATGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAG 9677

Qy 266 PheSerSerPheAlaAlaAspSerLeuArgLysProPheArgSerArgArgLeu 285
Db 9676 TTCTCCAGCTCTTCGCCATCGACAGCATCTCGCGCAAGCCCTTCGCGACGCCCGCCTC 9617

Qy 286 ArgAspThrAlaProGlyTyrThrLeuGlnTyrGlyAlaAlaProCysProLeuPro 305
Db 9616 AGGACACGCGCGCGGAGCAGCGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCG 9557

Qy 306 AlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyr 325
Db 9556 CGGTTCGCGCGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTAC 9497

Qy 326 GlyAlaGlyGluProAlaArgLeuGlyValaArgGluAlaGluValProThrAlaPro 345
Db 9496 CGCGCGGCGAGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9437

Qy 346 ProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAla 365
Db 9436 CCGCTCTCTGTTCACCTCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9377

Qy 366 AlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeu 385
Db 9376 CGCGCGGCGCGCGCGCGCTGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9317

Qy 386 ValArgArgProGlyProHisLeuSerTyrProValGluThrLeuLeuAla 402
Db 9316 GTCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9266

```

RESULT 13

AC012350/c
LOCUS
DEFINITION

AC012350 178168 bp DNA linear HTG 01-APR-2000
Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered
pieces.

AC012350
VERSION
KEYWORDS
SOURCE
ORGANISM

AC012350.3 GI:7381803
HTG; HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178168)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-16N9
Unpublished
2 (bases 1 to 178168)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguski,W., Boukhalil,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collins,A.,
Cooke,E., DeRubeis,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hags,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McWan,P., McMurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:6479001.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3593
Center clone name: 16.N.9
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172070 bases at least Q40
Consensus quality: 174479 bases at least Q30
Consensus quality: 175597 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 178668; sum-of-contigs
Quality coverage: 6.4 in Q20 bases; agarose-fp
Quality coverage: 6.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1694: contig of 1694 bp in length
* 1695 1794: gap of 100 bp
* 1795 4437: contig of 2643 bp in length
* 4438 4537: gap of 100 bp
* 4538 6644: contig of 2107 bp in length
* 6645 6744: gap of 100 bp
* 6745 10092: contig of 3348 bp in length
* 10093 10192: gap of 100 bp

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:7144942.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3560
Center clone name: 16_J_9

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely for
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 944: contig of 944 bp in length
* 945: gap of 100 bp
* 1045: contig of 947 bp in length
* 1992: gap of 100 bp
* 2091: contig of 916 bp in length
* 2092: gap of 100 bp
* 3008: contig of 930 bp in length
* 3108: gap of 100 bp
* 4037: contig of 930 bp in length
* 4038: gap of 100 bp
* 4138: contig of 919 bp in length
* 5057: gap of 100 bp
* 5157: contig of 923 bp in length
* 6080: gap of 100 bp
* 6180: contig of 924 bp in length
* 7104: gap of 100 bp
* 7204: contig of 933 bp in length
* 8136: gap of 100 bp
* 8237: contig of 930 bp in length
* 9167: gap of 100 bp
* 9267: contig of 901 bp in length
* 10168: gap of 100 bp
* 10268: contig of 945 bp in length
* 11212: gap of 100 bp
* 11312: contig of 938 bp in length
* 12250: gap of 100 bp
* 12351: contig of 927 bp in length
* 13277: gap of 100 bp
* 13378: contig of 928 bp in length
* 14276: gap of 100 bp
* 14376: contig of 906 bp in length
* 15281: gap of 100 bp
* 15382: contig of 918 bp in length
* 16300: gap of 100 bp
* 16400: contig of 942 bp in length
* 17342: gap of 100 bp
* 17442: contig of 918 bp in length
* 18360: gap of 100 bp
* 18459: contig of 937 bp in length
* 19397: gap of 100 bp
* 19497: contig of 961 bp in length
* 20458: gap of 100 bp
* 20557: gap of 100 bp
* 20558: contig of 925 bp in length
* 21482: gap of 100 bp
* 21582: contig of 933 bp in length
* 22515: gap of 100 bp
* 22615: contig of 899 bp in length
* 23515: gap of 100 bp
* 23615: contig of 916 bp in length
* 24530: gap of 100 bp
* 24531: contig of 923 bp in length
* 24631: gap of 100 bp
* 25554: contig of 942 bp in length
* 25554: gap of 100 bp
* 26595: contig of 932 bp in length
* 26596: gap of 100 bp
* 26627: contig of 932 bp in length
* 27272: gap of 100 bp
* 27628: contig of 933 bp in length
* 28661: gap of 100 bp
* 28661: contig of 932 bp in length
* 29693: gap of 100 bp
* 29792: contig of 916 bp in length
* 29793: gap of 100 bp
* 30708: contig of 920 bp in length
* 30709: gap of 100 bp
* 30809: contig of 920 bp in length
* 31729: gap of 100 bp
* 31729: contig of 912 bp in length
* 32741: gap of 100 bp
* 32741: contig of 921 bp in length
* 33761: gap of 100 bp
* 33762: contig of 943 bp in length
* 33862: gap of 100 bp
* 34805: contig of 931 bp in length
* 34905: gap of 100 bp
* 35336: contig of 947 bp in length
* 35936: gap of 100 bp
* 36882: contig of 929 bp in length
* 36883: gap of 100 bp
* 37911: contig of 925 bp in length
* 37912: gap of 100 bp
* 38012: contig of 925 bp in length
* 38336: gap of 100 bp
* 39036: contig of 954 bp in length
* 39337: gap of 100 bp
* 39990: contig of 914 bp in length
* 40090: gap of 100 bp
* 40091: contig of 914 bp in length
* 41004: gap of 100 bp
* 41004: contig of 914 bp in length
* 42018: gap of 100 bp
* 42018: contig of 897 bp in length
* 43015: gap of 100 bp
* 43115: contig of 937 bp in length
* 44052: gap of 100 bp
* 44052: contig of 907 bp in length
* 44152: gap of 100 bp
* 45059: contig of 933 bp in length
* 45159: gap of 100 bp
* 46092: contig of 918 bp in length
* 46192: gap of 100 bp
* 47110: contig of 928 bp in length
* 47111: gap of 100 bp
* 48238: contig of 919 bp in length
* 48238: gap of 100 bp
* 49157: contig of 933 bp in length
* 49257: gap of 100 bp
* 50190: contig of 929 bp in length
* 50290: gap of 100 bp
* 51219: contig of 929 bp in length
* 51319: gap of 100 bp
* 52248: contig of 929 bp in length
* 52348: gap of 100 bp
* 52449: contig of 934 bp in length
* 53249: gap of 100 bp
* 53382: contig of 943 bp in length
* 53382: gap of 100 bp
* 54325: contig of 901 bp in length
* 54326: gap of 100 bp
* 54426: contig of 911 bp in length
* 55327: gap of 100 bp
* 55327: contig of 911 bp in length
* 56337: gap of 100 bp
* 56337: contig of 919 bp in length
* 57356: gap of 100 bp
* 57357: contig of 960 bp in length
* 58416: gap of 100 bp
* 58416: contig of 960 bp in length

DEFINITION Zebrafish DNA sequence from clone DKEY-100F21 in linkage group 20, complete sequence.

ACCESSION BX510653

VERSION BX510653.8 GI:37051088

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 62848)

AUTHORS Beasley,H.

TITLE Direct Submission

JOURNAL Submitted (29-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk

COMMENT On Sep 29, 2003 this sequence version replaced gi:35761991.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhihong Bao and Sean Eddy, submitted), and those beginning 'arr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml

DKEY-100F21 is from a Zebrafish BAC library

VECTOR: pIndigoBAC-5.

FEATURES

source

1. 62848

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clones="DKEY-100F21"

/clone_lib="DanioKey"

ORIGIN

Alignment Scores:

Pred. No.:	6,09e-10	Length:	62848
Score:	685.50	Matches:	169
Percent Similarity:	95.03%	Conservative:	28
Best Local Similarity:	47.21%	Mismatches:	74
Query Match:	32.29%	Indels:	87
DB:	5	Gaps:	13

US-10-087-080-32 (1-402) x BX510653 (1-62848)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20

Db	55678	ATGAGTTGGAGTTTC-----TCTGGCAGTCGCTTGTGGACAAGCCGCTGGACCTG	55625
Qy	21	GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu	40
Db	55624	TGCAGTCACATGCGTGAACGTCGCTTCACCGCTGTCACG-----GAGGAGGAATTG	55571
Qy	41	GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln	60
Db	55570	GCCTCGGATGGAGACTGCTAGCCAAC-----	55544
Qy	61	GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAla	80
Db	55543	-----AGTCCAGAGCCCGCTCCCT-----	55523
Qy	81	AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyProGlyAlaGly	100
Db	55522	-----GTTCCGACGGG-----	55511
Qy	101	GlyAlaGlySerGlyGlyAlaArgSerLysProTyrThrArgArgProLysProPro	120
Db	55510	-----AAGCGAAACCATATACGCTAGACCCAAACCACT	55475
Qy	121	TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr	140
Db	55474	TACTCTTACATTCGACCTATTGCAATGCCATACCGATTCCACACCGCTCGCTTACA	55415
Qy	141	LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr	160
Db	55414	CTTGCTGAATAAAGCAATACCTTAATGAAAAATATCCCGTTTTTCGGGCGAGCTACAG	55355
Qy	161	GlyTrpArgAnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu	180
Db	55354	GGGTGGCGAAACTCTGACGCGATACTGTGCTTAACAGCACTCTTTTAAAGATTG	55295
Qy	181	ArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu	200
Db	55294	CGAGATCCCTCGCTCCGTGGGCGAAGACAATATTGGATTGTAACCCACACAGGAG	55235
Qy	201	TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro	220
Db	55234	TACACTTTTGACAGCGCGTTTTCCGCGGAGGAGGAGCGATCAGT-----	55187
Qy	221	ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProPro	240
Db	55186	-----AAGAGATCTCGGCGCGGGAATCCCGAGAGCGC	55151
Qy	241	AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla	260
Db	55150	GTGCCCGCGGATGACAGC-----AGGCTTCAGCGCGCGAGCGAAGC-----	55109
Qy	261	SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe	280
Db	55108	-----GTCCTTAAGTTTTCCAGCTCTTCGCCATAGACATCTCTCAGCAACCGTTC	55055
Qy	281	-----ArgSerArgArgLeuArgAspThrAlaProGlyThrThr-----LeuGlnTrpGly	297
Db	55054	ATGCGAGCGGACGACAGCAGATGACACTTGTGTTGTTGTACACACAGCTAATGCTGTC	54995
Qy	298	AlaAlaPro-----CysProProLeuProAlaPheProAla	309
Db	54994	GCTGCTCTCTCACTTTCTGCCAGTTGCCATGGGCTATCCGCCACAGAGAGATTCCAGTC	54935
Qy	310	LeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGlu	329
Db	54934	-----AGTCTGAGGTGTCGAATGCTTTCCAAATCTACAGGTGTATAACCGCGAC	54884
Qy	330	ProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaProProLeu	347
Db	54883	ATTTCATAACATC---AGTCGAATGGCGACATACAGAGTTCCCTTCGGGGTTTG	54833

Search completed: March 31, 2004, 08:01:03

Job time : 4503 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 04:42:59 ; Search time 444 Seconds

(without alignments)
3846.342 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVPRAHGDKQSDL.....AALVRPGPHLSYPVETLLA 402

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 337963 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-O/cn2_1/USPTO.spcol_p/US10087080/runat_29032004.114945.24363/app_query.fasta_1.583
-DB=N_Geneseq_23Jan04 -QFNT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080 @CGN 1.1.885 @runat_29032004.114945.24363 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	100.0	1209	6 ABQ81557	Abq81557 Gene up-r
2	1412.5	66.5	1760	7 ABT42112	Abt42112 Toxicity
3	790	37.2	658	6 ABQ39023	Abq39023 Oligonuc1
C 4	790	37.2	658	6 ABQ39022	Abq39022 Oligonuc1
5	488.5	23.0	2487	9 AD875311	Ad875311 Prostate
6	488.5	23.0	5181	6 ABL68935	Ab168935 Kidney ca
7	487.5	22.0	2271	6 AB235021	Ab235021 Human gen
C 8	483	22.8	658	6 ABQ39025	Abq39025 Oligonuc1

9	483	22.8	658	6 ABQ39024	Abq39024 Oligonuc1
10	472.5	22.3	3946	2 AA231671	Aa231671 Human FKX
11	472.5	22.3	3946	2 AA238079	Aa238079 Human for
12	466.5	22.0	1976	2 AA228095	Aa228095 Human FRE
13	466.5	22.0	2187	7 AB234783	Ab234783 Coding se
14	459	21.6	3482	7 ABX76288	Abx76288 Lung canc
15	446.5	21.0	1965	9 ADB31305	Adb31305 Testoster
16	445	21.0	1355	9 AB234708	Ab234708 Coding se
17	442.5	20.8	1380	9 ADD29558	Add29558 Human tum
18	429	20.2	1662	2 AA231672	Aa231672 Human FKX
19	429	20.2	1662	2 AA238080	Aa238080 Human for
20	419.5	19.8	2745	7 AB220402	Ab220402 Oncocfoet
21	419.5	19.8	2753	6 ABT03399	Abt03399 Ovary cel
22	416	19.6	3786	5 AA272683	Aa272683 DNA encod
23	411	19.4	2559	9 ADD15216	Add15216 Human ser
24	410	19.3	3038	5 ABV25148	Abv25148 Human pro
25	407.5	19.2	2934	6 AB199306	Ab199306 Mouse isc
26	406.5	19.1	2512	5 AA27345	Aa27345 DNA encod
27	406	19.1	1422	6 ABK15142	Abk15142 Human hep
28	406	19.1	2872	6 ABL69589	Ab169589 Prostate
29	406	19.1	2872	6 ABT11015	Abt11015 Human bre
30	406	19.1	2872	7 ACC50184	Acc50184 Breast ca
31	406	19.1	11613	6 ABK15141	Abk15141 Human hep
32	405	19.1	3283	4 AA26570	Aa26570 Human win
33	405	19.1	6458	6 ABK48984	Abk48984 Genomic D
34	404.5	19.1	1860	2 AAQ50632	AaQ50632 Brain fac
35	404.5	19.1	2832	2 AAQ50631	AaQ50631 Brain fac
36	403.5	19.0	1309	4 AAF27658	Aaf27658 DNA encod
37	403	19.0	2712	6 ABK48987	Abk48987 CDNA enco
38	403	19.0	6021	6 ABK48986	Abk48986 Genomic D
39	398	18.7	2448	9 ADD29815	Add29815 Human tum
40	395.5	18.6	2106	2 AA28103	Aa28103 Freatcl1 g
41	391	18.4	1287	5 AAS72684	Aas72684 DNA encod
42	391	18.4	4454	4 AAK90985	Aak90985 Human dig
43	391	18.4	4454	5 AAS32020	Aas32020 Human liv
44	391	18.4	4454	6 ABN90375	Abn90375 Human liv
45	387	18.2	4450	4 AAK90986	Aak90986 Human dig

ALIGNMENTS

RESULT 1

ABQ81557
ID ABQ81557 standard; cDNA; 1209 BP.

XX AC ABQ81557;

XX DT 30-DEC-2002 (first entry)

XX DE Gene up-regulated in metastatic colorectal cancer.

XX KW Colorectal cancer; metastasis; differential expression; cytostatic;
XX diagnosis; gene therapy; vaccine; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..1209

XX FT /*tag= a

XX FT /product= "Metastatic colon cancer polypeptide"

XX PN WO200268677-A2.

XX PD 06-SEP-2002.

XX PF 27-FEB-2002; 2002WO-US006001.

XX PR 27-FEB-2001; 2001US-027206P.

XX PR 02-APR-2001; 2001US-0281149P.

XX PR 17-APR-2001; 2001US-0284555P.

XX XX (EOSB-) BOS BIOTECHNOLOGY INC.

XX PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX Mack DH, Markowitz SD;
 XX WPI; 2002-698677/75.
 DR P-PSDB; ABP54691.
 XX
 XX New genes that are up- or down-regulated in colorectal cancer, useful for
 PT diagnosing colorectal cancer in a subject, or for identifying modulators
 PT of colorectal cancer-associated proteins and genes for treating
 PT colorectal cancer.
 XX
 XX Claim 5; Page 252; 260pp; English.
 XX
 XX The present sequence is the nucleotide sequence of a human gene that
 CC exhibits increased expression in metastatic colorectal cancer (MCC)
 CC samples. The gene is up-regulated in colon cancer-derived metastases
 CC compared to normal colon tissue. It is an example of claimed nucleic acid
 CC molecules that are up- or down-regulated in metastatic colorectal cancer
 CC cells. Such MCC-associated nucleic acids are useful in diagnostic and
 CC prognostic applications, in screening applications e.g. biochips, for
 CC identification of variant MCC-associated sequences, in informatics, for
 CC expression of MCC-associated proteins, in drug screening assays for
 CC identification of modulators of MCC-associated proteins, which are useful
 CC for treating this cancer, in gene therapy, as DNA vaccines, and as
 CC antisense or ribozyme modulators of MCC
 XX
 XX Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 7,176-69 Length: 1209
 Score: 2123.00 Matches: 402
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-087-080-32 (1-402) x ABQ81557 (1-1209)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 Db 1 ATGAAGTTGAGGTGTTGCTCCCTCGCGGCCCAAGCAGGGGACAGGAGGCGAGTACCTG 60
 Qy 21 GLuGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu 40
 Db 61 GAGGCGCGGGCGGCGGCGGCGGCTCCCTGCTGCGCGGCGGAGGAGCTCCCTG 120
 Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGln 60
 Db 121 GGCTCAGATGGGACTCGCGGCGCAAGCCCTCGCGGGCGGCGGCGGCGGCGGCGGCGG 180
 Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluAlaAlaAlaAla 80
 Db 181 GCGACGGCGACAGAGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
 Db 241 GCTGCTCAGCGGTGTTGGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
 Db 301 GCGCGGGGAGCGGCGGAGGTTGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 360
 Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeuThr 140
 Db 361 TACTCGTACATCGGCTCATCGCCATGCCATCGGACTCGGCGGCGGCGGCGGCTTGACG 420
 Qy 141 LeuAlaGluIleAenGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160
 Db 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGCGGCGGCGGCGGCGG 480
 Qy 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
 Db 481 GGCTGGCGCACTCCGTGGCGGCCCAACACTTTGCTCAACGACTGCTTCGTCAAGGTGCTG 540

Qy 181 ArgAspProSerArgProTTPGlyAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200
 Db 541 CGGACCCCTCGCGGCTTGGGCGAGGACAACTACTGGATGCTCAACCCCAACAGGAG 600
 Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysSerHisArgAlaPro 220
 Db 601 TACACCTTCGCGGACGGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCG 660
 Qy 221 ValProAlaProGlyLeuArgProGluAlaProGlyLeuProAlaAlaAlaProProPro 240
 Db 661 GTCCCGCGCGCGGCTGCGGCGCGAGAGGCGCGGCGCTCCCGCGCGCGCGCGCGCC 720
 Qy 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaAla 260
 Db 721 GCGCGCGCGCGCGGCTCGCGCGATCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 Qy 261 SerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
 Db 781 AGCCCGCGGCGCAAGTCTTCAGCTCTCTCCCATTCAGACATCTTCGCGCAAGCCCTTC 840
 Qy 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTTPGlyAlaAlaPro 300
 Db 841 CGCAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCTTCAGTGGGCGCGCGCGCGCG 900
 Qy 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320
 Db 901 TGCCCGCGCGCTCGCGGCTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCTG 960
 Qy 321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaAlaVal 340
 Db 961 CGGCTCGCGGTACGGTG 1020
 Qy 341 ProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaAlaProAlaLysPro 360
 Db 1021 CCACGCGCGCGCGCGCGCTCTCTGCTTCACCTCTCCCGCGCGCGCGCGCGCGCGCG 1080
 Qy 361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380
 Db 1081 CTCGAGGCGAGCC 1140
 Qy 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400
 Db 1141 CTGAGGCGCGCGCTTAGTCCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
 Qy 401 LeuAla 402
 Db 1201 CTAGCT 1206

RESULT 2
 ID ABT42112 standard; DNA; 1760 BP.
 XX AC ABT42112;
 XX DT 26-JUN-2003 (first entry)
 XX DE Toxicity modelling related rat gene SEQ ID No 1814.
 XX KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 XX database; drug screening; toxicity assay; rat; ds.
 XX OS Rattus norvegicus.
 XX PN WO200295000-A2.
 XX PD 28-NOV-2002.
 XX PF 22-MAY-2002; 2002WO-US016173.
 XX PR 22-MAY-2001; 2001US-0292335P.
 XX PR 13-JUN-2001; 2001US-0297523P.
 XX PR 19-JUN-2001; 2001US-0298925P.

PR 10-JUL-2001; 2001US-0303807P.
 PR 10-JUL-2001; 2001US-0303808P.
 PR 10-JUL-2001; 2001US-0303810P.
 PR 28-AUG-2001; 2001US-0315047P.
 PR 27-SEP-2001; 2001US-0324928P.
 PR 22-OCT-2001; 2001US-0330462P.
 PR 01-NOV-2001; 2001US-0330867P.
 PR 21-NOV-2001; 2001US-0331805P.
 PR 19-DEC-2001; 2001US-0336144P.
 PR 16-DEC-2001; 2001US-0340873P.
 PR 21-FEB-2002; 2002US-0357842P.
 PR 21-FEB-2002; 2002US-0357843P.
 PR 21-FEB-2002; 2002US-0357844P.
 PR 15-MAR-2002; 2002US-0364134P.
 PR 08-APR-2002; 2002US-0370144P.
 PR 08-APR-2002; 2002US-0370206P.
 PR 08-APR-2002; 2002US-0370247P.
 PR 17-APR-2002; 2002US-0372794P.
 PR 21-APR-2002; 2002US-0371679P.
 XX
 PA (GENE-) GENE LOGIC INC.
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 XX MPI; 2003-148464/14.
 DR
 XX
 PT Predicting at least one toxic effect of a compound, useful for toxicity
 PT modeling, comprises preparing a gene expression profile of a tissue or
 PT cell sample exposed to the compound, and comparing the gene expression
 PT profile to a database.
 XX
 PS Example 4; Page; 446pp; English.
 XX
 CC The invention relates to a novel method of predicting at least one toxic
 CC effect of a compound. The method comprises a gene expression profile of a
 CC tissue or cell sample exposed to the compound, and comparing the gene
 CC expression profile to a database comprising at least part of the data or
 CC information given in the specification. The methods are useful for
 CC predicting at least one toxic effect of a compound, predicting the
 CC progression of a toxic effect of a compound, predicting the renal
 CC toxicity of a compound, or identifying toxicity markers in tissues or
 CC cells exposed to known renal toxin. The genes are useful as toxicity
 CC markers in drug screening and toxicity assays, in monitoring disease or
 CC physiological states, or disease progression. This polynucleotide
 CC represents a rat DNA sequence relating to the toxic effect database
 CC described in the specification. NOTE: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the World Intellectual Property
 CC Organization
 XX
 SQ Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,05e-43 Length: 1760
 Score: 1412.50 Matches: 300
 Percent Similarity: 77.64% Conservative: 16
 Best Local Similarity: 73.71% Mismatches: 78
 Query Match: 66.53% Indels: 17
 DB: 7 Gaps: 6

US-10-087-080-32 (1-402) x AET42112 (1-1760)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 Db 207 ATGAATTGGAGTG-TTCGTCCACGCGCAGC-CACGGGGACAAAATGGCAGCGATCTG 264
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
 Db 265 GAGGGGGGGGGCAGCAGCGAGTGTCCATCTCCGCGGTGTGAGACTCTCTTA 324
 QY 41 GlySerAspGlyAspCysAlaAla---LysProSerAlaGlyGlyAlaArgAspThr 59
 Db 325 GGCTCAGACGGGAGTGTGCAGCCAAACAGCCGGGGGGGGCAGCGGGGGGATCTG 384

QY 60 Gln---GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIlePro 78
 Db 385 GAAGGTGGGGGGGGAGAGGAATTCGAGTGGGGGGGAGCACCACAGAC----- 435
 QY 79 AlaAlaAlaAlaAlaValValAlaGluGlyAla-----GluAlaGlyAlaAlaGly 96
 Db 436 -----GATCCCGAGGTGACCGATGCGAGCAACGCGGCGGCGGCGGCGG 483
 QY 97 ProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaAlaArgSerLysProTyrThrArgArg 116
 Db 484 CCGTGGCGGGCAGCGTGGCGGGGTGAGGGCGGCGGCGGCGGCGGCGGCGGCGG 543
 QY 117 ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 136
 Db 544 CCGACGCCCTCTACTCTTACATCGCACTCATCGCCATGCGCATCGGCACCTCGCGGGC 603
 QY 137 GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArg 156
 Db 604 GGACGGCTGAGCTGGCCGAGATCAACGAGTACCTCATGGGCAAGTTCCCTTTCCGG 663
 QY 157 GlySerTyrThrGlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe 176
 Db 664 GCGAGCTACACGGGTGGGGCAACTCCGTGCGCCACACCTCTCGCTCAACGACTGTTTC 723
 QY 177 ValLysValLeuArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsn 196
 Db 724 GTCAAGGTGCTGGCGACCCCTCGCGGCCCTGGGCAAGGACATTAATTGATGTTCTCAAC 783
 QY 197 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSer 216
 Db 784 CCCAACAGCGAATACACCTTCGCGAGGGGTCTTCGCGCGCGCGCGCAAGCGCTCAGC 843
 QY 217 HisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
 Db 844 CACCGGACCACTGCTCCGCATCGGGCTACGGCGAGGAGGCCACCCGAGACTCGCGGG 903
 QY 237 AlaProProAlaProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGln 256
 Db 904 ACCCGCAGCGCGGGCCACCACCGCGCTCTCTCCCAATCGCGCGCTCGCGCGCTCGCCAG 963
 QY 257 GluGluArgAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeu 276
 Db 964 GAGGAGGGTCCAGCGCGCGGAGAGTTCCTCAGCTCTTCGCCATCGACAGCATCTTC 1023
 QY 277 ArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrp 296
 Db 1024 AGCAAGCCGTTTCGACGCGCGCGCA-CGCGACCCGCGCTCTGGGGGTGCGACTACCTGG 1082
 QY 297 GlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCys 316
 Db 1083 AGCGCTGCTCCCTGCGCGCGCTATCCGCGCTATCCGCGCTCTCTCCCGCTCGTCCGGC 1142
 QY 317 ArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArg 336
 Db 1143 GGTGCCCTGTCGCGCTCTGTGTTACGGCGC-GGCGAGCCACCGCTGTGGCGTGGCG 1201
 QY 337 GluAlaGluValProProThrAlaProProLeuLeuAlaProLeuProAlaAlaAla 356
 Db 1202 GGGCGCGAGGTGCGCGCGCGCGG---CCCTGTGTGTGGCGCGCTCTCCACCGCGCGC 1258
 QY 357 ProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeu 375
 Db 1259 CCAGCCAAAGCCATTTTCGAGGTCCGAGACCGCGCGCGCGCGCGCTGTACTGCCCTTA 1318
 QY 376 ArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyr 395
 Db 1319 CGGTGCCACGCGCTCGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTCTAC 1378
 QY 396 ProValGluThrLeuLeuAla 402
 Db 1379 CGGGTGGAGACGCTGCTAGCT 1399

QY 313 -AlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaAa 332
 Db 1179 CCGCGCCCTCCGCGCTCCGCGCCCAAGCGGCG-----GCCCGGGCGCTCAGCG 1232
 QY 332 gLeuGlyAlaArgGluAla----- 338
 Db 1233 CTGGCGCGCTCCGCTTCCTTCATCGAGAGCATCATCGGGCGAGCTTGGCGCGCGGCC 1292
 QY 339 ----GluValProProThraProProLeuLeuAlaProLeuProAlaAlaAaPr 357
 Db 1293 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCGCTCG-CCCTCGCGCTCGCC 1351
 QY 357 o-----AlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLe 375
 Db 1352 GGTGCGCGCGCGCGCGCGCTCCGATCCAGCGGAGGAGCG----- 1391
 QY 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeu-ValArgArgPro---GlyProHisLeuS 394
 Db 1392 -----TGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCACCCTCCCT 1441
 QY 394 eTyrPro 396
 Db 1442 CQTGCGCG 1449

RESULT B

ABQ39025/C

ID ABQ39025 standard; DNA; 658 BP.

XX AC ABQ39025;

XX DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25616.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX Homo sapiens.

XX WO200218632-A2.

XX PD 07-NAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:

CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

XX Sequence 658 BP; 269 A; 222 C; 108 G; 59 T; 0 U; 0 Other;

Alignment Scores: 6.42e-10 Length: 658
 Pred. No.: 483.00 Matches: 120
 Score: 60.89% Conservative: 79
 Best Local Similarity: 59.41% Mismatches: 79
 Query Match: 22.75% Indels: 0
 DB: 6 Gaps: 0

US-10-087-080-32 (1-402) x ABQ39025 (1-658)

QY 200 GluTyrThrPheAlaAspGlyValPheArgArgArgLeuSerHisArgAla 219
 Db 658 GAGTATATTTCGTCACGGGGTTTTTCGTCGTCGTAAGCGTTTTAGTTATCGCGG 599
 QY 220 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAaProPro 239
 Db 598 TCGGTTTCGCTTCGGTTTCGGTTTCGAGAGAGTTTCGGTTTTTCGCTTCGTCG 539
 QY 240 ProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 259
 Db 538 TTCGCTTCGTCGTTTCGGTTTTTCGTTTCGATCGCTTCGTTTCGTTTCGTTTCG 479
 QY 260 AlaSerProAlaGlyLysPheSerSerSerPheAlaAlaAspSerLeuLeuArgLeuPro 279
 Db 478 GTTAGTTTCGCGGGTAAGTTTTTTAGTTTTTCGTTTCGATCGATAGTATTTTGGTAAGTT 419
 QY 280 PheArgSerArgArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAla 299
 Db 418 TTTCGTAGTCGTCGTTTTCGGGATACGGTTTTTCGGGACGACGTTTTAGTGGGCGCTCGCG 359
 QY 300 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAaProCysArgAlaLeu 319
 Db 358 TTTTGTTCGTCGTTTCGCGTTTTTCGCGTTTTTTTCGCGCGGTTTTGTAGGGTTTTG 299
 QY 320 LeuProLeuCysAlaTyrGlyAlaGlyGluProAlaAaArgLeuGlyAlaArgGluAlaGlu 339
 Db 298 TTGTCGTTTTGCGGTACGCGCGGCGAGTCGCGCGGTTGCGCGCGCGCGGTCGAG 239
 QY 340 ValProProThraProProLeuLeuAlaProLeuProAlaAlaAlaProAlaLys 359
 Db 238 GTGTTATCGATCGCGTCGTTTTTTTGTGTTGTTTTCGCGCGGTTTTTCGTTAAG 179
 QY 360 ProLeuArgGlyProAlaAlaGlyAlaHisLeuTyrCysProLeuArgLeuProAla 379
 Db 178 TTATTTTCGAGGTTTCGCGCTCGCGCGCGGATTTGTTGTTTTCGCGGTTGTCGTA 119
 QY 380 AlaLeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThr 399
 Db 118 GTTTTGTAGCGGTTTTAGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCGTTTCG 59
 QY 400 LeuLeu 401
 Db 58 TTTTTA 53

RESULT 9

ABQ39024

ID ABQ39024 standard; DNA; 658 BP.

XX AC ABQ39024;

XX XX

CC This sequence encodes the human forkhead transcription factor gene, designated FKHL7, of the invention. FKHL7 can be used in a novel method for treating or preventing the development of a congenital heart disease (CHD) in a subject. The FKHL7 sequences can be used for diagnosis, prognosis, monitoring, prevention and treatment of CHD. They can also be used for the production of transgenic animals and drug screening.

XX
SQ Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.04e-09 Length: 3946
Score: 472.50 Matches: 165
Percent Similarity: 39.07% Mismatches: 44
Best Local Similarity: 30.94% Mismatches: 145
Query Match: 22.26% Indels: 182
DB: 2 Gaps: 22

US-10-087-080-32 (1-402) x AAZ31671 (1-3946)

QY 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
DB 287 CCGAGGCGCGGAGGAGCAGCCAGCGAGCGCGGAGGAGGCGGAGCGAGCGCGGAGC 346
QY 28 Ala-----ProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
DB 347 GCACAGCGCAGCGGCGCGGACCA---GCTCGCGCGGCGCGGACTCGGACTCGGCGGCC 403
QY 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGly 63
DB 404 GCGCGGCGCGCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
QY 64 GluGlnSerAlaGlyGlyProGlyAlaGluGlnAlaLysProAlaAlaAla----- 81
DB 464 GAGCGG-----GGGCCA---TGACAGGCGCGCTACTCGGTCTCCAGCGCCCAACT 508
QY 82 -----AlaAlaValValAlaGluGlyAlaGluAlaGly 93
DB 509 CCCTGGAGTGTGGCTACTCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 567
QY 94 AlaAlaGlyProGly-----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113
DB 568 GCGGCGGCGGCGGCGGTACACCGCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
QY 99 -----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113
DB 628 CACGCGGAGGAGTACCGCGGCGGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
QY 114 ThrArgArgPro-----LysProProTyrSerTyrIleAlaLeuIle 127
DB 676 ACGCGCAGCG 735
QY 128 AlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyr 147
DB 736 ACCATGGCCATCCAGACG 795
QY 148 LeuMetGlyLysPheProPhePheArgGlySerTyrThrGlyTyrArgAsnSerValArg 167
DB 796 ATCATGGACCGCTTCCCTTCTACCGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 855
QY 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTyr 187
DB 856 CACAACCTCTCGCTCAACAGAGTGTCTGTCAGGAGTGTGCGCGCGCGCGCGCGCGCGCGCG 912
QY 188 GlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207
DB 913 GCGAAGGCGAGCTACTGACCGCTGACCGCGACTCTTACACATGTTTCGAGAGCGGAGC 972
QY 208 PheArgArgArgLysArgLysSerHis----- 217
DB 973 TTCTGCGGCGCGCGCGCGCTTCAAGAGAGAGGAGCGCGGTGAAGGAGGAGGAGGAGGAGGAG 1032
QY 218 -----ArgAlaProValProAlaProGly 225

DB 1033 GACAGGTCGACCTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089
QY 226 LeuArgProGluGluAlaProGlyLeuProAlaAlaPro----- 238
DB 1090 -----CGGAGCAGCG 1143
QY 239 -----ProAlaProAlaAlaProAlaAla 246
DB 1144 CAGGACATCAAGACCGAGAACGGTAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1203
QY 247 Ser-----ProArgMetArgSerPro----- 253
DB 1204 GCGCGCTCGGAGCGGAGCG 1263
QY 254 ---AlaArgGlnGluArgAlaSerProAlaGlyLysPhe----- 266
DB 1264 AGCAGGAGCTGTCCAGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1323
QY 266 ----- 266
DB 1324 AGCCTGGAGCGGTGCGGATTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1383
QY 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278
DB 1384 CATAGCCAGGCTTTCAGCTGGACACATCATGCTCGTGGGGGGTGGCGCGCGAGCG 1443
QY 279 -----ProPheArgSerArgArgLeuArgAspThrAlaProGly 291
DB 1444 GCGGCGCGGAGCTCAGCTCCGCGCTTCTGGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1503
QY 291 YThrThrLeuGlnTyrGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311
DB 1504 ATCGACCG 1563
QY 311 uProAlaAlaProCysArgAlaLeuProLeuCysAlaTyrGlyAlaGlyGluProAla 331
DB 1564 CCCTGCG 1623
QY 331 aArgLeuGlyAlaArgGluAlaGluValProProThrAla-----ProProLeuLeu 348
DB 1624 GCGG---GGGCG 1680
QY 348 uLeuAlaProLeu-----ProAlaAlaAlaProAlaAla 359
DB 1681 TAGCG 1740
QY 359 sProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysPro-----Le 375
DB 1741 GCGGTGGACAACCG 1776
QY 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgPro 389
DB 1777 CCGGTACCGAGCGCA-----CCTGCTGCTGCC 1804
RESULT 11
AAZ38079
ID AAZ38079 standard; DNA; 3946 BP.
XX
XX AAZ38079;
AC
XX
XX 22-FEB-2000 (first entry)
DT
DE Human forkhead transcription factor gene, FKHL7 DNA sequence.
XX
XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;
XX transgenic animal; drug screening; ss.
OS Homo sapiens.
XX
XX W0953060-A2.
XX
XX 21-OCT-1999.
XX

PF 14-APR-1999; 99WO-US008148.
 PR 15-APR-1998; 98US-0081870P.
 PR 22-MAY-1998; 98US-0083352.
 XX (TOWA) UNIV IOWA RES FOUND.
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;
 PI WPI; 1999-620429/53.
 DR P-PSDB; AAY55721.
 DR New isolated human forkhead transcription factor gene, FKHL7, used to,
 XX e.g. develop products for the diagnosis.
 XX Claim 1; Fig 1; 99pp; English.
 XX The invention provides a human forkhead transcription factor gene, FKHL7.
 CC The FKHL7 protein can be produced by standard recombinant methodology.
 CC The products can be used for diagnosis, prognosis, monitoring, prevention
 CC and treatment of glaucoma. They can also be used for the production of
 CC transgenic animals and drug screening. The present sequence represents
 CC the DNA sequence of the FKHL7 gene
 XX
 SQ Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;

Alignment Scores:
 Pred. NO.: 6.04e-09 Length: 3946
 Score: 472.50 Matches: 165
 Percent Similarity: 39.07% Conservative: 44
 Best Local Similarity: 30.84% Mismatches: 145
 Query Match: 22.26% Indels: 182
 DB: 2 Gaps: 22

US-10-087-080-32 (1-402) x AA238079 (1-3946)

QY 8 ProAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
 DB 287 CCGAGCGCGGAGGAGCGAGCCAGCGCGCGGAGCGCGGAGCGCGGAGCGCGGAG 346
 QY 28 Ala-----ProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
 DB 347 GCACAGCGCAGCGCGCGGACCA---GCTCGCGCGCGCGCGCGCGCGCGCGCGCG 403
 QY 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyAlaAlaArgAspThrGlnGlyAspGly 63
 DB 404 GCG 463
 QY 64 GluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaAlaProAlaAlaAla----- 81
 DB 464 GAGCGG-----GGGCCA---TGCAGCGCGCGCTACTCGGTGTCCAGCGCCCAACT 508
 QY 82 -----AlaAlaAlaValValAlaGluGlyAlaGluAlaGly 93
 DB 509 CCCTGGGAGTGGTCCCTACCTCGCGCGGAGAGAGCTACTACCGC-GCGCGCGCGCG 567
 QY 94 AlaAlaGlyProGly-----AlaAlaGlyProGly----- 98
 DB 568 GCGCGCGCGCGCGGTACACCGCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
 QY 99 -----AlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyr 113
 DB 628 CACGCGGAGAGTACCGCGCGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
 QY 114 ThrArgArgPro-----LysProProTyrSerTyrIleAlaLeuIle 127
 DB 676 AGCGCGCAGCGCGCAGCAGCATGTGTGAAGCGCGCGCTATAGCTACATCGCGCTCATC 735
 QY 128 AlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyr 147
 DB 736 ACCATGGCATCTCAGAACGCGCGGACAAAGAGATACCCCTGAACGCGCATCTACCATTC 795
 QY 148 LeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrPArgAsnSerValArg 167

RESULT 12
 AA228095

DB 796 ATCATGGACCGCTTCCCTTCTACCGGACAAACAGGAGGCTGGCAGAACGATCCGC 855
 QY 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTrp 187
 DB 856 CACAACCTCTCGCTCAACGAGTCTTGGTCAAGTGGCGGCGAGCAAGAAGCGC--- 912
 QY 188 GlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207
 DB 913 GGCAGGCGCAGCTACTGGACGCTGGACCGCGGACTCTTACACATGTTTCGAGAACGGCAGC 972
 QY 208 PheArgArgArgArgLysArgLysSerHis----- 217
 DB 973 TTCCTGGCGCGCGCGCGCTTCAAGAGAAGAGCGGTGAAGGACAAAGAGAGAGAAG- 1032
 QY 218 -----ArgAlaProValProAlaProGly 225
 DB 1033 GACAGGCTGCACCTCAAGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1089
 QY 226 LeuArgProGluGluAlaProGlyLeuProAlaPro----- 238
 DB 1090 -----CCGGAGAGCGCGCGCGCAACCGCGCGCGTCCGACGCGCGCGCGCGCGCATC 1143
 QY 239 -----ProProAlaProAlaAlaProAla 246
 DB 1144 CAGGACATCAAGACGAGACCGTACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1203
 QY 247 Ser-----ProArgMetArgSerPro----- 253
 DB 1204 GCGCGCGCTGGCGAGCG 1263
 QY 254 ---AlaArgGlnGluGluArgAlaSerProAlaGlyLysPhe----- 266
 DB 1264 AGCAGCAGCCTGTCGCGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1323
 QY 266 ----- 266
 DB 1324 AGCTTGGAGCGTGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1383
 QY 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278
 DB 1384 CATACCGAGGGCTTCAGCGTGCACCAATCATGACGTGCTCGCGGGGTGCGCGCAGAGC 1443
 QY 279 -----ProPheArgSerArgLeuArgAspThrAlaProGly 291
 DB 1444 GCGCGCGCGGAGCTCAGCTCGCGCGCTTCTGGCGCTCGCGCGCGCGCGCGCGCG 1503
 QY 291 YThrThrLeuGlnTrpGlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLe 311
 DB 1504 ATCGCACCCCGCTGGCGCTCGCGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCG 1563
 QY 311 uProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAl 331
 DB 1564 CCCTGACGCGACCTCCAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1623
 QY 331 aArgLeuGlyAlaArgGluAlaGluValProThrAla-----ProProLeuLe 348
 DB 1624 GCGG---GGGCG 1680
 QY 348 uLeuAlaProLeu-----ProAlaAlaAlaProAlaLys 359
 DB 1681 TACGCGCGCGCGAGCG 1740
 QY 359 sProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysPro-----Le 375
 DB 1741 GCGGTGGCAACACCC---TGCGGACTACTCTCTCGCT 1776
 QY 375 uArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgPro 389
 DB 1777 CCGTCCACGAGAGCA-----GCTCGTGTGCGC 1804

ID AAZ28095 standard; cDNA, 1976 BP.
 AC AAZ28095;
 XX
 XX
 DT 31-JAN-2000 (first entry)
 XX
 XX Human FREAC3 protein encoding cDNA.
 XX
 XX FREAC3 gene; eye disease; eye developmental defect; therapy; mutation;
 KW glaucoma; anterior segment dysgenesis; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 159..1820
 FT CDS /tag= a
 FT /product= "FREAC3"
 FT 403
 FT mutation /tag= b
 FT /note= "a G to C transversion mutation at this position
 FT can result in a Ser82Thr mutation in helix 1 of the
 FT FREAC3 forkhead domain"
 FT 419
 FT mutation /tag= c
 FT /note= "a G to C transversion mutation at this position
 FT can result in a Ile87Met mutation in helix 1 of the
 FT FREAC3 forkhead domain"
 FT
 XX WO954493-A2.
 PN
 XX
 XX 28-OCT-1999.
 PD
 XX
 XX 16-APR-1999; 99WO-IB001024.
 PF
 XX
 XX 17-APR-1998; 98US-0082206P.
 PR
 XX 08-MAY-1998; 98US-0084784P.
 PR
 XX (UYAL-) UNIV ALBERTA.
 PA
 XX
 XX Walter MA, Jordan T, Raymond V;
 PI
 XX
 XX WPI; 1999-634007/54.
 DR
 XX P-PSDB; AAY41277.
 DR
 XX
 XX Use of mutations in the FREAC3 gene for detecting predisposition to eye
 PT diseases or developmental defects and for developing products for
 PT prevention and therapy.
 PT
 XX
 XX Claim 50; Fig 2; 65pp; English.
 PS
 XX
 XX The invention relates to the use of mutations in the FREAC3 gene for
 CC detecting predisposition to diseases of the eye or developmental defects
 CC and for developing products for prevention and therapy. The method
 CC comprises analyzing nucleic acid of the mammal to determine whether the
 CC nucleic acid contains a mutation in a FREAC3 gene, where the presence of
 CC the mutation is an indication that the mammal has an increased likelihood
 CC of developing a disease of the eye. The methods can be used for the
 CC diagnosis of increased likelihood of developing a disease of the eye,
 CC e.g. glaucoma, or a developmental defect. The methods can also be used
 CC for the prevention and treatment of diseases of the eye. Transgenic
 CC animals containing the mutated FREAC3 gene can be used to screen for
 CC therapeutic compounds that modulate anterior segment dysgenesis or the
 CC onset or progression of glaucoma via a FREAC3-dependent or FREAC3-
 CC affected pathway. The present sequence represents the cDNA encoding the
 CC FREAC3 polypeptide
 XX
 XX Sequence 1976 BP; 362 A; 736 C; 658 G; 220 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 5,86e-09 Length: 1976
 Score: 466.50 Matches: 164
 Percent Similarity: 39.24% Conservative: 42
 Best Local Similarity: 31.24% Mismatches: 133

Query Match: 21.97% Indels: 187
 DB: 2 Gaps: 22
 US-10-087-080-32 (1-402) x AAZ28095 (1-1976)
 QY 22 GlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeu-Gl 41
 DB 6 GGGAGAGGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 65
 QY 41 YSerAspGlyAspCysAlaAla-----LysProSerAlaGlyGlyAl 56
 DB 66 CCGGGACTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
 QY 56 aArgAspThrGlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAlaGluAl 76
 DB 126 GCGGGCGGCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 170
 QY 76 aileProAlaAlaAla-----AlaAlaAlaValva 86
 DB 171 TACTCGGTGTCCAGGCCCAACTCCTGGGAGTGGTGCCTTACCTCGGCGCGGAGCAGC 230
 QY 86 lAlaGluGlyAlaGluAlaGlyAlaGlyProGly----- 98
 DB 231 TACTACGCG-GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 289
 QY 99 -----AlaGlyGlyAlaGlySerGlyGl 106
 DB 290 GAGCGTGTACTCGACCCCTCGCACCGCGAGCAGTACC CGGCGCGCATGCCCGCGCCTA 349
 QY 106 uGlyAlaArgSerLysProTyrThrArgArgPro-----LysProPr 120
 DB 350 CGGG-----CCCTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397
 QY 120 oTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuTh 140
 DB 398 CTATAGCTACATCGGCTCATCACCATGGCCATCCAGACGCCCGCGCGCGCGCGCG 457
 QY 140 rLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrTh 160
 DB 458 CCTGACGCGCATCTACCAAGTTCATCATGACCGCTTCCCTTCTACCGGACACACAGCA 517
 QY 160 rGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheVallysValLe 180
 DB 518 GGGCTGCGAAGACAGCATCCGCCACACCTCTCGCTCAAGAGTGTCTCGTCAAGGTGCC 577
 QY 180 uArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGl 200
 DB 578 GCGCGACGACACAGAGGCGG--GGCAAGGCGAGCTACTGGACGCTGGACCGCGACTCCTA 634
 QY 200 uTyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysSerHisArg---- 218
 DB 635 CAACATGTTTCGAGAACGCGCAGCTTCTCGGCGCGCGCGCGCGCGCGCGCGCG 694
 QY 218 ----- 218
 DB 695 GTTGAAGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 754
 QY 219 -AlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaLap 238
 DB 755 CAGCCCCCGCGCGCGCGG-----CGGAGAGCGCGCGCGCGCGCGCGCGCG 805
 QY 238 o-----Pr 239
 DB 806 GCAGCG 865
 QY 239 oProAlaProAlaProAlaSer-----ProAr 249
 DB 866 GCGCCAGCG 925
 QY 249 gMetArgSerPro-----AlaArgGlnGluAlaArgAlaSerProAlaGlyLy 265
 DB 926 GATCGAGAGCG 985

Qy 265 sPhe----- 266
Db 986 CTGCGCTCGCGCGCGCTGAGCTGAGCGTGGATTCGCGCGCGCGCGCGCG 1045
Qy 267 -----SerSerSerPheAlaAlaAspSerLeuLeu-ArgL 278
Db 1046 GCGCTCG 1105
Qy 278 ys-----ProPheArgSerArgA 284
Db 1106 GCTGCG 1165
Qy 284 rgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProProl 304
Db 1166 GCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1225
Qy 304 euProAlaPheProAlaLeuLeuProAlaAla-----ProCysArgAlaLeuProL 322
Db 1226 CCAGAGCTCCCTACAGCTCCCTCGACCGACCTCCAGCGCGCGCGCGCG 1285
Qy 322 euCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProp 342
Db 1286 CG 1336
Qy 342 roThrAla-----ProProLeuLeuAlaProLeu----- 352
Db 1337 CCACTGCAACCTGCAAGCCATGAGCCTGTACGCGCGCGCGCGCGCGCG 1396
Qy 353 -----ProAlaAlaProAlaAlaProAlaArgGlyProAlaAlaAlaGlyAlaH 370
Db 1397 GGGCG 1440
Qy 370 isLeuTyrcysPro-----LeuArgLeuProAlaAlaLeuGlnAlaLeuV 386
Db 1441 -----TGCGCGACTACTCTCTGCTCGCTCGGTCCAGCAGCA-----G 1477
Qy 386 alaArgArgPro 389
Db 1478 CTCGTGCTGCC 1488
RESULT 13
ID ABZ34783
XX ABZ34783 standard; cDNA; 2187 BP.
XX AC ABZ34783;
XX DT 04-FEB-2003 (first entry)
XX DE Coding sequence SEQ ID 141, upregulated in osteogenesis.
XX KW Osteopathic; osteogenesis modulator; gene therapy; osteogenesis;
XX KW osteoporosis; bone disease; downregulator; human; singal transduction;
XX KW ss.
XX OS Homo sapiens.
XX PN WO20021745-A2.
XX PD 17-OCT-2002.
XX PF 05-APR-2002; 2002WO-IB002211.
XX PR 05-APR-2001; 2001US-0281400P.
XX PA (AVET) AVENTIS PHARMA SA.
XX Garcia T, Roman Roman S, Baron R, Call K, Theilhaber J;
PI Connolly T, Jackson A, Bushnell SE, Rawadi G;
XX WPI; 2003-058567/05.
XX Novel isolated nucleic acid upregulated/downregulated in osteogenesis,
PT useful for bone disease therapy in subject.

XX Claim 26; Page 153-154; 237pp; English.
XX The present invention relates to novel nucleotide sequences, which are
CC differentially expressed in models of osteogenesis upon being put in
CC contact with a stimulator of osteogenesis. The present sequence is one
CC such sequence. This sequence can be used for diagnosing osteoporosis/bone
CC disease in a patient, promoting osteogenesis and/or preventing
CC osteoporosis/bone disease. The present sequence encodes a protein
CC involved in intracellular signal transduction
XX SQ Sequence 2187 BP; 411 A; 736 C; 615 G; 425 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6.33e-09 Length: 2187
Score: 466.50 Matches: 157
Percent Similarity: 45.35% Conservative: 48
Best Local Similarity: 34.73% Mismatches: 159
Query Match: 21.97% Indels: 90
DB: Gaps: 18
US-10-087-080-32 (1-402) x ABZ34783 (1-2187)
Qy 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
Db 79 CGCGCTCGCGCGCGCGCTCGCTCCCGGCTCCAGATGACACCGAGCGCGGG---CCG 135
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySer---AspGlyAspCys 46
Db 136 CGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
Qy 47 AlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGlyGluGlnSer 66
Db 196 ATGACCG 255
Qy 67 AlaGlyGlyProGlyAlaGluAlaLarleProAlaAlaAlaAlaAlaValVal 86
Db 255 TCCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 315
Qy 87 AlaGluGlyAlaGluAlaAlaGlyAlaGlyProGlyAlaGlyGlyAlaGlySerGlyGlu 106
Db 316 GCGCGCTCGCTCGCTCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
Qy 107 GlyAlaArgSerLysProTyThrArgArgPro---LysProProTyThrSerTyThrLea 125
Db 376 GCGAAGAGCGAGCTCG 435
Qy 126 LeuAlaMetAlaAlaArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluLeuAsn 145
Db 436 CTCATCGTCATCGCCATCCAGAGCTCGCCAGCAGCGCTGACCGCTCAGCGAGATCTAC 495
Qy 146 GluTyLeuMetGlyLysPheProPheArgGlySerTyThrGlyTyThrArgAsnSer 165
Db 496 CAGTTCCTGCGCGCGCTTCCTCTTCGCGCGCGCTTACCGAGGCTGGGAAGACTCG 555
Qy 166 ValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArg 185
Db 556 GTGCGCCCAATCTCTCGCTCAACGAGTCTTCATCAGCTGCTTAAGGGCTCGCGCG 615
Qy 186 ProTrpGlyLysAspAsnTyTrpMetLeuAsnProAsnSerGluTyThrPheAlaAsp 205
Db 616 CCC---GGCAAGCGCCACTTACTGGACCTACCGACCGCGCGCGCGCGCGCGCGAG 672
Qy 206 GlyValPheArgArgArgLysArgLeuSerHisArgAlaPro----- 220
Db 673 GGCTCGTTCG 731
Qy 221 ValPro-----AlaProGlyLeuArg 227
Db 732 GTACCAACCGCGTGTGAGCGGCTTGGGCTTCGGGCGTTCGCTGCTGCCCGCGGCTTCA 791
Qy 228 ProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSer 247

Db 792 -----CTTCAGGGGCCCCCGTCCGGCGCGCTCGGCTCCACAGCCA 833
 Qy 248 ProArgMetArgSerPro-----AlaArgGln 256
 Db 834 GGGCGGCTACGGCGGCTCGACATGATCCCGCGGCTACGACGCGCGGGCGGCC 893
 Qy 257 GluGluArgAla-----SerProAlaGlyLysPheSer 268
 Db 894 CAGCCAGCGGCGCCCTCACCACACCACCACGCTCCGCGCATGTGCGCCAAACC 953
 Qy 269 SerPheAlaAlaAspSerIleLeu-----ArgLysProPhe 280
 Db 954 GGGTTCACCTACATGCGCCAGCTGCGCGTCCCGCGGACCCCGGGCGTCTCGTGGCG 1013
 Qy 281 ArgSerArgArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTTPGlyAlaAlaPr 300
 Db 1014 CCGGGCGCGCGCGCGGCGGCGTACGGCGCGACAGCAGC-----AGCAGGCC 1051
 Qy 300 cCysProProLeuProAlaPhe-----ProAlaLeuLeuPr 312
 Db 1062 GTTACCTCGTCCCGGCCATGCGGAGCGCCATGCAATGCCACTCGCCCTACACGAGCC 1121
 Qy 312 alalala-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTy 325
 Db 1122 TCGCGCGCACTGGAGCTCGCCCTGG-CGCTCGCCTTACCTCAAGCAGCGCGCTGCGCTGA 1180
 Qy 325 rGlyAlaGlyGlyProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaPr 345
 Db 1181 CCGCCAGCAGCAACCCCGCGCTGGCAGG-----CTGCACTCCAGCA 1225
 Qy 345 oProLeuLeuAlaProLeuProAlaAlaProAlaProAlaProLeuArgGlyProAl 365
 Db 1226 TGTCTCTCTACTCGCTGGAGCAGAGCTACTTCACACCAAGCGCTCG---CGAGGACCTCT 1282
 Qy 365 aAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLe 385
 Db 1283 CAGTGGGAGTGGC-----CCGTACGAGCATCACTCTACTCCAGTGT 1324
 Qy 385 u---ValArgArgProGlyProHisLeuSerTyr 395
 Db 1325 GTGACAGAAAGATTTCGTCTCAACTCAATGG 1358
 RESULT 14
 ABX76288
 ID ABX76288 standard; DNA; 3482 BP.
 XX AC ABX76288;
 XX DT 02-APR-2003 (first entry)
 XX DE Lung cancer-associated polynucleotide #152.
 XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
 XX KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 XX KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX OS Unidentified.
 XX PN WO200286443-A2.
 XX PD 31-OCT-2002.
 XX PF 18-APR-2002; 2002WO-US012476.
 XX PR 18-APR-2001; 2001US-0284770P.
 XX PR 10-MAY-2001; 2001US-0290492P.
 XX PR 09-NOV-2001; 2001US-0339245P.
 XX PR 13-NOV-2001; 2001US-0350666P.
 XX PR 29-NOV-2001; 2001US-0334370P.
 XX PR 12-APR-2002; 2002US-0372246P.

(EOSB-) EOS BIOTECHNOLOGY INC.
 PA Aziz N, Murray R;
 XX WPI; 2003-093161/08.
 XX P-PSDB; ABUS6559.
 *XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 PS Claim 22; Page 303-304; 453pp; English.
 XX
 CC The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridizes
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hyperensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
 CC invention
 XX
 SQ Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;
 Alignment Scores: Length: 3482
 Pred. No.: 1.69e-08 Matches: 157
 Score: 459.00 Conservative: 34
 Percent Similarity: 48.72% Mismatches: 142
 Best Local Similarity: 40.05% Indels: 63
 Query Match: 21.62% Gaps: 18
 DB: 7
 US-10-087-080-32 (1-402) x ABX76288 (1-3482)
 Qy 30 SerProLeuSerAlaAlaGly-----AspAspSerLeuGlySer--- 42
 Db 504 GCGCCAGCTGGCGCGCTCGGCTCTCCGGCCAGCCCGCGGACGATCCCTGAGCTCTCG 563
 Qy 43 AspGlyAspCysAlaAlaLysProSerAlaGlyGly-----GlyAlaArgAspThrGln 60
 Db 564 CAGAAGGGCGAGCGTCCGTTCGGGGAGCGCCAGGCGCGCCCGCCCGCCAGACGCGC 623
 Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGly-----AlaGluGluAlaIle 77
 Db 624 GGGGATCCAGAGCCCGGGGTGGCGAGCGCCCGCGCATGCTGCGGAGAGCGGGCGCGC 683
 Qy 78 ProAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGly-----AlaAla 95
 Db 684 GCCCGCGCAGCGGAGGTGCTGGCTACCGTGAAGGAA--GAGCGCGCGAGACGCGAGCA 741
 Qy 96 GlyProGlyAlaGlyGlyAlaGlySerGlyGluGlyAla-----ArgSerLysPro 112
 Db 742 GGGCGCGGTCCAGGGGAGGCCACGGCGCGGGCGGGCGGGCGCGCGCAAGCGC 801
 Qy 113 TyrThrArgArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArg 132
 Db 802 CCCCTGACGCGCGGAAGCGCCCTACGATACATCGCGTCTATCGCCATGCGCATCGCG 861
 Qy 133 AspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPhe 152
 Db 862 CAGCGCGCGAGCGCGCTCACGCTGGGCGGATCTACAGTTCATCACCAGCGCGCTTC 921


```
Db 337 CGCCACATCTCTCGCTCAAGAGTGTTCATCAAGCTGCTAAGGCCCTCGGGGGCCC 396
Qy 187 TrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGly 206
Db 397 ---GGCAAGGGCCACTACTGACCATCGACCGGCCAGCGAGTTCATGTTTCGAGGAGGC 453
Qy 207 ValPheArgArgArgLysArgLeuSerHisArgAlaPro-----Val 221
Db 454 TCGTTCGCGCCCGGCC--GCGCGCTTCAGGCGGAAGTGCAGGCGCTCAAGCCCATGTA 512
Qy 222 Pro-----AlaProGlyLeuArgPro 228
Db 513 CCACGCGTGTGTAGCGGCTTGGGCTTCGGGGCTCGCTGCTGCCCGCCAGGGCTTCGA-- 569
Qy 229 GluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSerPro 248
Db 570 -----CITCCAGGCGCCCGCTCGCGCGCTCGCTGCTGCCACAGCCAGGG 614
Qy 249 ArgMetArgSerPro-----AlaArgGlnGlu 257
Db 615 CGGCTACGGCGGCTCGACATGATGCCCGGGGTACGACCGCGGGCGGGCCCCCAG 674
Qy 258 GluArgAla-----SerProAlaGlyLysPheSerSerSer 269
Db 675 CCAGCGGCACCCCTCACCACCACCACCACCACCGCTCCCGCACATGTCGCCCAACCGGG 734
Qy 270 PheAlaIleAspSerIleLeu-----ArgLysProPheArg 281
Db 735 TTCACCTACATGGCCAGCTCCCGGTGCCCGGGACCGGGGGCTCGGTGCGGCGG 794
Qy 282 SerArgArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCy 301
Db 795 GGGCGGCGGGCGCGGCGGACTACGGCGCGGACAGCAGC-----AGCAGCCCGGT 842
Qy 301 sProProLeuProAlaPhe-----ProAlaLeuLeuProAl 313
Db 843 ACCCTCGTCCCGGCCATGGCGAGGCCATCGAATGCCACTCGGCCCTACAGAGCCCTGC 902
Qy 313 aAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTyrG1 326
Db 903 GGGCGACTGGAGTCGCTCG--CGCCTCGCCTTACCTCAAGCAGCGCCTGCCCTGACGC 961
Qy 326 yAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaProPr 346
Db 962 CCAGCAGCAACCCCGCCCTCGGCGG-----CCTGCACCTCCAGCATGT 1006
Qy 346 oLeuLeuAlaAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAl 366
Db 1007 CCTCTACTCTGGAGCAGAGTACTTGACACAGACGCTCG---CGAGGACCTCTCAG 1063
Qy 366 aGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeu-- 385
Db 1064 TGGGACTGCC-----CCGTATCAGCATCACTACTACTCTACTCTGTTGTG 1105
Qy 386 -ValArgArgProGlyProHisLeuSerTyr 395
Db 1106 ACAGAAAAGATTTCGTCCTCAACTCAATGG 1136
```

Search completed: March 31, 2004, 06:46:05
Job time : 467 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 05:56:45 ; Search time 100 Seconds
(without alignments)
2230.903 Million cell updates/sec

Title: US-10-087-080-32
Perfect score: 2123
Sequence: 1 MKLEVFVRAHGXQGSGL.....AALVRRPGHSLVPEVTELLA 402

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10087080/runat_29032004_114947_24501/app_query.fasta_1.583
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080@cgn_1_1_103@runat_29032004_114947_24501 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	22.1	320	4	US-09-833-381-1266
2	468.5	22.1	3946	3	US-09-083-351-1
3	468.5	22.1	3946	3	US-09-083-352-1
4	446.5	21.0	1965	4	US-09-220-133-21
5	425	20.0	1659	3	US-09-083-351-3
6	425	20.0	1659	3	US-09-083-352-3
7	404.5	19.1	1860	2	US-08-331-644-3
8	404.5	19.1	1860	5	PCT-US93-04102-3
9	403.5	19.0	1309	4	US-09-576-594-927
10	400.5	18.9	2830	1	US-07-882-292-1
11	400.5	18.9	2830	2	US-08-331-644-1
12	400.5	18.9	2830	5	PCT-US93-04102-1

13	360.5	17.0	1155	2	US-08-331-644-4	Sequence 4, Appli
14	360.5	17.0	1155	5	PCT-US93-04102-4	Sequence 4, Appli
15	324.5	15.3	5080	4	US-09-976-594-495	Sequence 495, App
16	309.5	14.6	1605	4	US-09-087-134-10	Sequence 10, Appl
17	307	14.5	1634	4	US-09-087-134-13	Sequence 13, Appl
18	306.5	14.4	1793	3	US-09-113-309-1	Sequence 1, Appli
19	306.5	14.4	1793	3	US-09-521-109-1	Sequence 1, Appli
20	306.5	14.4	1793	4	US-09-562-332-1	Sequence 16, Appl
21	297	14.0	1668	4	US-09-087-134-16	Sequence 1, Appli
22	264	12.4	2517	1	US-07-906-930E-1	Sequence 1136, Ap
23	264	12.4	3465	1	US-09-023-655-1136	Sequence 3, Appli
24	263	12.4	2939	1	US-07-906-930E-3	Sequence 314, App
25	251.5	11.8	3394	4	US-09-620-312D-314	Sequence 3, Appli
26	236	11.1	3342	2	US-08-742-753-3	Sequence 1, Appli
27	233.5	11.0	3441	2	US-08-742-753-1	Sequence 1, Appli
28	231	10.9	441529	3	US-09-103-840A-1	Sequence 2, Appli
29	229.5	10.8	4403765	3	US-09-103-840A-2	Sequence 1, Appli
30	228.5	10.8	2888	4	US-08-765-907A-1	Sequence 1, Appli
31	226	10.6	441529	3	US-09-103-840A-1	Sequence 2, Appli
32	224	10.6	4403765	3	US-09-103-840A-2	Sequence 3, Appli
33	216	10.2	1869	4	US-09-372-668-3	Sequence 3, Appli
34	216	10.2	1869	4	US-09-697-377-3	Sequence 3, Appli
35	216	10.2	1869	4	US-09-696-868-3	Sequence 3, Appli
36	216	10.2	1869	4	US-09-697-341-3	Sequence 3, Appli
37	211.5	10.0	12001	1	US-08-458-568A-11	Sequence 35, Appl
38	210.5	9.9	2481	4	US-09-894-998A-35	Sequence 50, Appl
39	210.5	9.9	12425	4	US-09-616-289-50	Sequence 1, Appli
40	208	9.8	8438	1	US-07-945-283-1	Sequence 1, Appli
41	207.5	9.8	2150	2	US-08-318-837-1	Sequence 48, Appl
42	206.5	9.7	2561	4	US-09-616-289-48	Sequence 153, App
43	205.5	9.7	2109	4	US-09-370-838-153	Sequence 1, Appli
44	204	9.6	34094	4	US-09-292-034-1	Sequence 11, Appl
45	203	9.6	12001	1	US-08-458-568A-11	

ALIGNMENTS

RESULT 1

US-09-833-381-1266
; Sequence 1266, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1266
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(320)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1266

Alignment Scores:
Pred. No.: 1.09e-17
Score: 470.00
Percent Similarity: 89.62%
Best Local Similarity: 88.68%
Query Match: 22.14%
DB: 4
Length: 320
Matches: 94
Conservative: 1
Mismatches: 9
Indels: 2
Gaps: 0

US-10-087-080-32 (1-402) x US-09-833-381-1266 (1-320)

Qy 190 AspaanTyrTrpMetLeuAnProAsnSerGluTyrThrPheAlaaspGlyValPheArg 209

Db 2 GACAACTACTGGATGCTCAACCCACAGAGAGTACACCTTCGCCGACGGGTCTTCGC 61
Qy 210 ArgArgArgLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGlu 229
Db 62 CGCGCGCGAAGCGCTCAACCAACCGCGCGGTCCCGCGCGCGGTGGCGGCGCGAG 121
Qy 230 GluAlaProGlyLeuProAlaProProAlaProProAlaProAlaProAlaSerProAr 249
Db 122 GAGGCGCGCGCTTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181
Qy 249 gMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSe 269
Db 182 CATGCGCTCG 241
Qy 269 rPheAla-11eAspSerLleuArgLysProPheArgSerArgArgLeuArgPThrA 289
Db 242 CTTTCATCATCTACAGATCTTCGCAATCTTCGCGATCGCGCGCGCGCGCGCG 301
Qy 289 laProGlyThrThr 293
Db 302 CCGCGGAGGACG 315

RESULT 2

US-09-083-351-1
; Sequence 1, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 475..2133
US-09-083-351-1
Alignment Scores:
Pred. No.: 1.87e-16 Length: 3946

Score: 468.50 Matches: 165
Percent Similarity: 38.92% Conservative: 44
Best Local Similarity: 30.73% Mismatches: 143
Query Match: 22.07% Indels: 186
DB: 3 Gaps: 23
US-10-087-080-32 (1-402) x US-09-083-351-1 (1-3946)
Qy 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGlnGlyAlaGlyGlySerAsp 27
Db 287 CCGAGGCGCGGAGAGAGCCAGCCAGCGAGCGCGCGGAGAGCGGAGCGGAGCGGAGC 346
Qy 28 Ala-----ProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAsp 43
Db 347 GCACAGCGCAGCGCGCGCGGACCA---GCTCGCGCGCGCGCGGACTCGGACTCGCGCG 403
Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyCysAlaArgAspThrGlnGlyAspGly 63
Db 404 GCGCGCGCGCGCGCGCGCGCGAGCGAGGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 463
Qy 64 GluGlnSerAlaGlyGlyProGlyAlaGluAlaLleProAlaAlaAla-----81
Db 464 GAGCGG-----GGGCGCA---TGCAGGCGCGCTACTCGGTCTCCAGCGCCCACT 508
Qy 82 -----AlaAlaAlaValAlaAlaGluGlyAlaGluAlaGly 93
Db 509 CCCTGGGAGTGTGCCCTACTCCTCGCGCGGAGAGAGTACTACCGG-GCGGCGCGCG 567
Qy 94 AlaAlaGlyProGly-----98
Db 568 GCGCGCGCGCGCGCTACACCGCATCGCGCGCGCGCGCGCTACCGG-----CCCTAC 627
Qy 99 -----AlaGlyGlyAlaGlySerGlyGlyAlaArgSerLysProTyr 113
Db 628 CACGCCGAGCAGTACCGCGCGCGCATGGCGCGCGCTACCGG-----CCCTAC 675
Qy 114 ThrArgArgPro-----LysProProTyrSerTyrLleAlaLle 127
Db 676 ACGCCGCGCGCGAGCCCAAGACATGGTGAAGCGCGCTATAGCTACATCGCGCTCATC 735
Qy 128 AlaMetAlaLleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluLleAsnGly 147
Db 736 ACCATGGCCATCCAGAACGCCCGGCAAGAGATACCTCTGAACGCGCATCTACCACTTC 795
Qy 148 LeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAnSerValArg 167
Db 796 ATCATGGACGGCTTCCCTTCTACCGGCAACAAGCAGCGGCTGGCAGAACACATCCGC 855
Qy 168 HisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSerArgProTyr 187
Db 856 CACAACCTCTCGCTCAACGAGTGTCTTCAAGTGGCGCGCGCGCGCGCGCGCGCG---912
Qy 188 GlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyVal 207
Db 913 GGCAAGGCGCAGTACTGGACGCTGGACCGCGACTCCTTACACATGTTCGAGAACGCGCAG 972
Qy 208 PheArgArgArgLysArgLeuSerHis-----217
Db 973 TTCTCGCGCGCGCGCGCGCTTCAAGAGAGAGCGCGGTGAAGCAAGGAGGAGAG 1032
Qy 218 -----ArgAlaProValProAlaProGly 225
Db 1033 GACAGGCTGCACCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG---1089
Qy 226 LeuArgProGlnGluAlaProGlyLeuProAlaAlaPro-----238
Db 1090 -----CCGAGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1143
Qy 239 -----ProProAlaProAlaAlaProAla 246
Db 1144 CAGGACATCAAGACCGAGAGCGGTACGTGCCCTCGCGCGCGCGCGCGCGCGCGCG 1203
Qy 247 Ser-----ProArgMetArgSerPro-----253


```
Db 913 GGCAGGCGAGCTACTGGACCTGGACCGGAGCTCTCAACATGTTTCGAGACGCGACG 972
Qy 208 PheArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 217
Db 973 TTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1032
Qy 218 -----ArgAlaProValProAlaProGly 225
Db 1033 GACAGGCTGACCTCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1089
Qy 226 LeuArgProGluGluAlaProGlyLeuProAlaPro----- 238
Db 1090 -----CCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1143
Qy 239 -----ProProAlaProAlaProAlaProAla 246
Db 1144 CAGGACATCAAGACCGAGAACGTTAGTGCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1203
Qy 247 Ser-----ProArgMetArgSerPro----- 253
Db 1204 GCGCGCTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1263
Qy 254 ---AlaArgGlnGluArgAlaSerProAlaGlyLysPhe----- 266
Db 1264 AGCAGCAGCTGTCCAGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1323
Qy 266 ----- 266
Db 1324 AGCTGACGGTGGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1383
Qy 267 ---SerSerSerPheAlaIleAspSerIleLeu-ArgLys----- 278
Db 1384 CATAGCAGGCGCTTACGCTGGACACATCATGCTGCGGGGGTGCCTGCGCGCAGAGC 1443
Qy 279 -----ProPheA-gSerArgArgLeuArgAspThrAlaProGly 291
Db 1444 GCGCGCGCGAGCTACGCTCGCGCTTCTGGCTTCGGCGCGCGCGCGCGCGCGCGCGCG 1503
Qy 291 YThrThrLeuGlnTrpGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311
Db 1504 ATCGCACCCCGCTGGCTGGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1563
Qy 311 uProAlaAla-----ProCysArgAlaLeuLeuProLeuCysAlaTyrclyAlaGly 329
Db 1564 CCCTGACAGCAGACCTCCAGCGGGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1623
Qy 329 uProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAla-----ProPr 346
Db 1624 G-----CGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1674
Qy 346 oLeuLeuAlaProLeu-----ProAlaAlaPro 357
Db 1675 AGCTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1734
Qy 357 oAlaAlaProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrcysPro----- 374
Db 1735 GGCTCGCGCGTGGACACCCCG-----TGCGCGCTACTCT 1770
Qy 375 ----LeuArgLeuProAlaAlaLeuGlnAlaLeuValArgPro 389
Db 1771 CTGCTCGGCTCACAGCAGCA-----GCTCGTCTGCTCC 1804

RESULT 4
US-09-220-132-21
; Sequence 21, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shvian, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE OF INVENTION: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
```

```
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1965
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-220-132-21

Alignment Scores:
Pred. No.: 1,31e-15 Length: 1965
Score: 446.50 Matches: 150
Percent Similarity: 44.55% Conservative: 42
Best Local Similarity: 34.80% Mismatches: 121
Query Match: 21.03% Indels: 120
DB: 4 Gaps: 19

US-10-087-080-32 (1-402) x US-09-220-132-21 (1-1965)
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAspGlyAspCysAla 47
Db 13 GCCCGGAGACACCTCTCTCTCC-----TCGTCGCTCTCTCCGCTCTCTCGGCC 63
Qy 48 AlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGlyGluGlnSerAla 67
Db 64 TCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 90
Qy 68 GlyGlyGlyProGlyAlaGluGluAlaIleProAlaAlaAlaAlaAlaValAlaAla 87
Db 91 -----AGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 114
Qy 88 GluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyAlaGlySerGlyGluGly 107
Db 115 -----AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156
Qy 108 AlaArgSerLysProTyThrArgArgPro---LysProProTyThrSerTyThrAlaLeu 126
Db 157 AAGAGGAGGAGCTCGGGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 216
Qy 127 IleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGlu 146
Db 217 ATCGTCATGGCCTCCAGAGCTCCCGCAGCAGCGCTGACGCTCAGCGAGATCTACAG 276
Qy 147 TyrLeuMetGlyLysPheProPheArgGlySerTyThrGlyTyThrArgAsnSerVal 166
Db 277 TTCCTGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
Qy 167 ArgHisAsnLeuSerLeuAsnAspCysPheValValLeuArgAspProSerArgPro 186
Db 337 CGCCCAATCTCTCTCTCAACAGAGTCTTCAACAGTCTTCAACAGTCTTCAACAGTCTT 396
Qy 187 TrpGlyLysAspAsnTyThrMetLeuAsnProAsnSerGluTyThrPheAlaAspGly 206
Db 397 ---GSCAAGGGCCACTACTGGACCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 453
Qy 207 ValPheArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 221
Db 454 TCGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Qy 222 Pro----- 228
Db 513 CCACCGCGTGTGAGCGGCTTGGGCTTCGGGCGCTCGGCTCGGCTCGGCTCGGCTCGGCT 569
Qy 229 GluGluAlaProGlyLeuProAlaAlaProProProAlaProAlaAlaProAlaSerPro 248
Db 570 -----CTTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 614
Qy 249 ArgMetArgSerPro-----AlaArgGlnGlu 257
```


QY 259 ArgAlaSerProAlaGlyLysPhe-----266
Db : : : : :
808 GGGAGCAGACCCCGGGCAGCGCTCGCGCGCGCGCGCTCAGCCTGAGCGGTGCGAT 867
QY 267 -----SerSerSerPheAla 271
Db : : : : :
868 TCCGGCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 927
QY 272 lleAepSerileLeu-ArgLys-----278
Db : : : : :
928 GTGGACAACATCATGCTGCTCGCGGGGTGCGCGAGAGCGCGCGCGCGAGCTCAGC 987
QY 279 ---ProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGI 297
Db TCCGGCGCTTCGGCGCTCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1047
QY 297 yAlaAlaProCysProProLeuProAlaPheProAlaLeuProAlaAla-----Pr 315
Db : : : : :
1048 CTCGGCGCTACTCGCGCGCGCGAGCTCCCTCTACAGCTCCCGCTGACGACCTCC 1107
QY 315 cCysArgAlaLeuLeuProLeuCysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAl 335
Db AGCGCGGCGAGCTCG 1158
QY 335 aArgGluAlaGluValProProThra-----ProProLeuLeuAlaProLe 352
Db GCGGGCGCGCGCGCGCGCTTACCATGCAACTGCAAGCCATGAGCTGTACGGCGCGCG 1218
QY 352 u-----ProAlaAlaAlaProAlaLysProLeuArgGI 363
Db GAGCG 1219
QY 363 yProAlaAlaGlyAlaHisLeuTyCysPro-----LeuArgLeuProAl 379
Db CCCC-----TCCCGACTACTCTCTGCTCGCTCCGTCACGAGC 1314
QY 379 aAlaLeuGluAlaAlaLeuValArgArgPro 389
Db 1315 AGCA-----GCTCGCTCGTCC 1330

RESULT 6

US-09-083-352-3
; Sequence 3, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, ROAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1659 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-083-352-3

Alignment Scores:

Pred. No.:	1,51e-14	Length:	1659
Score:	425.00	Matches:	140
Percent Similarity:	39.91%	Conservative:	40
Best Local Similarity:	31.04%	Mismatches:	112
Query Match:	20.02%	Indels:	159
DB:	3	Gaps:	19

US-10-087-080-32 (1-402) x US-09-083-352-3 (1-1659)

QY	72	GlyAlaGluGluAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlu	91
Db	58	GGCGCGGAGAGAGAGTACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	111
QY	92	AlaGlyAlaAlaGlyProGlyAla-----	99
Db	112	ACCGCGCATCG	171
QY	100	GlyGlyAlaGlySerGlyGluGlyAlaAlaArgSerLysProTyThrArgArgPro	117
Db	172	GGCGCATCG	219
QY	118	-----LysProProTyThrSerTyThrLeuAlaAlaAlaAlaAlaAlaAlaAlaAla	133
Db	220	AAGGACATGGTGAAGCG	279
QY	134	SerAlaGlyGlyArgLeuThrLeuAlaGluLeuAlaAlaAlaAlaAlaAlaAlaAla	153
Db	280	GGCGCGGAGAGAGATACCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	339
QY	154	PhePheArgGlySerTyThrGlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsn	173
Db	340	TTTACCGGAGACCAAGCAGCGGTGGCAGAACAGCATCCGCCACACCTCTCGCTCAAC	399
QY	174	AspCysPheValLysValLeuArgAspProSerArgProTyGlyLysAspAsnTyTrp	193
Db	400	GAGTCTTCGTCAAGGTCCCGCGCGAGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCG	456
QY	194	MetLeuAsnProAsnSerGluTyThrPheAlaAspGlyValPheArgArgArgLys	213
Db	457	ACGCTGGACCGGACTCTCAACATGTTTCGAGAACGCGAGCTTCTCTCGCGCGCGCG	516
QY	214	ArgLeuSerHis-----	217
Db	517	CGCTTCAAGAGAGAGCGCGGTGAAGCAAGAGGAGAGAGAGAGAGAGAGAGAGAGAG	576
QY	218	-----ArgAlaProValProAlaProGlyLeuArgProGluGluAla	231
Db	577	GAGCG	627
QY	232	ProGlyLeuProAlaAlaPro-----	238
Db	628	GACGGCAACGCGCGCGCGGTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	687
QY	239	-----ProAlaProAlaAlaProAlaSer-----	247
Db	688	AACGGTAGTCCCTCG	747
QY	248	-----ProArgMetArgSerPro-----	258
Db	748	AGCGCGCGCGCGGTGCGCGCAAGATCGAGAGCGCGCGCGCGCGCGCGCGCGCGCG	807

Db 953 CAACCGG---GGCAGGCGCACTACTGGAGCGCTCGACCCCGAGTCCGCAGATATGTTGGA 1009
 Qy 204 aAspGlyValPheArgArgArgGlySerHisArgAlaProValProAlaPr 224
 Db 1010 CAACGGCAGCTTCCTGGCGCCGCAAGCGGTTCAAG---CGCAGCGCTACTCGATCC 1066
 Qy 224 oGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAl 244
 Db 1067 CGCCTG----- 1073
 Qy 244 aProAlaSerProArgMetArgSerProAlaArgGlnGluGluArgAlaSerProAlaGl 264
 Db 1074 -----GGGACTCTGCACCAAGGACAGCGCTGCCAGTGTGGA 1111
 Qy 264 yLyAspSerSer-SerPheAlaIleAspSerIleLeuArgLysProPheArgSerArgA 284
 Db 1112 GAATTTACTCTAGGATTT-----CCATTGTAGACGCTGTAGCGCGGG 1162
 Qy 284 rgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGly-----AlaAlaProCysP 302
 Db 1163 AGAGCGAAGGTAGGACTCCCGCTTCTTCTCCGATGGGGGTGGTTTCTGTCGCC 1222
 Qy 302 roProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuPro- 321
 Db 1223 CTCGGGTCTCCGAGACCCCGCGCCCGCC-CGGTTTTCGCCGCTTCGGATTCTTGACAC 1281
 Qy 322 -LeuCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArg 336
 Db 1282 ACTGTGTGGCGACAGCTGGGGCGCGCGAGCTTTAGCTCAGAGG 1327

RESULT 8

PCT-US93-04102-3
 ; Sequence 3, Application PC/TUS9304102
 ; GENERAL INFORMATION:
 ; APPLICANT: Tao, wufan
 ; APPLICANT: Lai, Eseng
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/04102
 ; FILING DATE: 19930430
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/882,292
 ; FILING DATE: 13-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-977-9550
 ; TELEFAX: 212-664-0525
 ; TELEX: 42523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1860 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA

PCT-US93-04102-3

Alignment Scores:
 Pred. No.: 2,09e-13 Length: 1860
 Score: 404.50 Matches: 133
 Percent Similarity: 40.95% Conservative: 30
 Best Local Similarity: 33.42% Mismatches: 132
 Query Match: 19.05% Indels: 104
 Ds: 5 Gaps: 13
 US-10-087-080-32 (1-402) x PCT-US93-04102-3 (1-1860)
 Qy 8 ProArgAlaAlaHis-----GlyAspLysGlnGlySerAspLeuGluGly 22
 Db 245 CGCGCGCTGCAGAGTGGCGTCTCGCTCCGGTCCGCCCTCCGGATCGCCTCGGGA 304
 Qy 23 AlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAsp----- 37
 Db 305 GGCAGGAGGCGCGGAGCGCGTGCCTATGTGCCCGCGCCACCCGCCGCCGCCAGC 364
 Qy 38 ---AspSerLeuGlySerAspGlyAspCys----- 46
 Db 365 TATGACCTGAGCAGGAGATGTCGATGCTCCGGCTCGGGAGGAGACAGACATCGA 424
 Qy 47 -----AlaAlaLysProSerAlaGlyGlyAlaArg-AspThrGlnGlyAspGl 63
 Db 425 CGTGTGGGGAGGCGGAGGACGACGAGGAGGAGGAGGATGACGACGAGGGCGGG 484
 Qy 63 yGluGlnSerAlaGlyGlyGly----- 70
 Db 485 CGCGCGCGCGCGCGCGGTCCCGTTCGCGAGTCCGCCCGCGCGCGGCGGCTCTTA 544
 Qy 70 ----- 70
 Db 545 CGCGCGGAGGTGCGATCTCGAGACCTGGAGGAGGAGGACGATGATGCTGCTCGC 604
 Qy 71 -----ProGlyAlaGluAlaIleProAlaAlaAlaAlaAlaValAlaGl 88
 Db 605 CCGCGCGCGCGCGCTCCCGCGCCTCCGGTCTCGGGTCTCGCGCGCGCGGACGGGTC 664
 Qy 88 uGlyAlaGluAlaGlyAla---AlaGlyProGlyAlaGly-----GlyAlaGlySe 104
 Db 665 GCGCGCTGCAGCGCGCGCGGAGCGGAGCGCGGCGGAGGTGTAGCGGCGCGGCGAC 724
 Qy 104 rGlyGluGlyAlaArgSerLysProTyThrArgArgProLysProTyProTySerTyTr 124
 Db 725 GCGCGCGCGCGCTAAGAT---CCGCTGGTG-----AAGCGCGCTACTCGTACAT 772
 Qy 124 eAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeuThrLeuAlaGluIl 144
 Db 773 CGCGCTCATCACATGGCCATCTCTCAGAGCGCCCAAGAGCGCTGACGCTCAGCGAGAT 832
 Qy 144 eAsnGluTyLeuMetGlyLysPheProPheArgGlySerTyThrGlyTyTrArgAs 164
 Db 833 CTGCGATTTCATCAGCAGCGCTTCCCTTACTACCGGAGAGTTCGCCGCTTGCGCAA 892
 Qy 164 nSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspProSe 184
 Db 893 CAGCATCGTGCACACCTGCTCGCTCAACAGCTGCTTCTCAAGATCCCGCGCAACCGG 952
 Qy 184 rArgProTrpGlyLysAspAsnTyTrpMetLeuAsnProAsnSerGluTyThrPheAl 204
 Db 953 CAACCCG---GGCAAGGGCAACTACTGACCGCTCGCGGAGGAGTTCCTGATGTTTGA 1009
 Qy 204 aAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaProValProAlaPr 224
 Db 1010 CAACGGCAGCTTCCTGGCGCCGCGCAAGCGGTTCAAG---CGCAGCGCTACTCGATCC 1066
 Qy 224 oGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAl 244
 Db 1067 CGCGCTG----- 1073
 Qy 244 aProAlaSerProArgMetArgSerProAlaArgGlnGluGluArgAlaSerProAlaGl 264

QY 375 uArgLeuProAlaLeu-----GlnAlaLeuValArgArgProGlyPr 391
 DB 1674 ACCT---CCGGACCGGTCAATGACTTCGACAGACGACGTCCTAGCGCCCGCGCG 1730
 QY 391 cHisLeuSerTyr 395
 DB 1731 CGTCTCTCTCTAC 1743

RESULT 12

PCT-US93-04102-1
 ; Sequence 1. Application PC/TUS9304102
 ; GENERAL INFORMATION:
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Lai, Seng
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/04102
 ; FILING DATE: 19930430
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/882,292
 ; FILING DATE: 13-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-977-9550
 ; TELEFAX: 212-664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2830 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHEetical: N
 ; ANTI-SENSE: N
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 443..1882
 ; OTHER INFORMATION:
 ; NAME/KEY: misc_feature
 ; LOCATION: 926..1255
 ; OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
 ; FEATURE:
 ; NAME/KEY: binding domain homology"
 ; LOCATION: 1883..1885
 ; OTHER INFORMATION: /note= "translation termination codon"
 PCT-US93-04102-1

Alignment Scores:
 Pred. No.: 5,328-13 Length: 2830
 Score: 400.50 Matches: 143
 Percent Similarity: 40.47% Conservative: 29
 Best Local Similarity: 33.65% Mismatches: 146

Query Match: 18.86% Indels: 109
 DB: 5 Gaps: 18
 US-10-087-080-32 (1-402) x PCT-US93-04102-1 (1-2830)
 QY 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
 DB 683 CCCCAGCGAGCGCGCGCGCCACGAGCGGACGACGACGAGGCG-----727
 QY 28 AlaProSerProLeu-----SerAlaAlaGlyAsp-----AspSer 39
 DB 728 ---CCCCAGCGCGCTTCTCTCCCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCA 784
 QY 40 LeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThr 59
 DB 785 CTTGGAGCCAAAGGCGAG-----CCAGCGCGCGCGCGCTCGGAGCTCGCGCC 832
 QY 60 GlnGlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaAlaProAla 79
 DB 833 GTCGGCGCGGACGAGGAGGAGGCGCGCGCGCGCTGGGGGGGAGGAG-----880
 QY 80 AlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAla 99
 DB 881 -----AAGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAG 907
 QY 100 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysPro 119
 DB 908 GACGGGAGGCG 961
 QY 120 ProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 139
 DB 962 CGTTTCACTCAACGCGCTCATCATGATGCGCATGAGGAGGAGGAGGAGGAGGAGGAG 1021
 QY 140 ThrLeuAlaGluIleAsnGluTyrLeuMetClyLysPheProPheArgGlySerTyr 159
 DB 1022 AGCTCAACGGGCACTACGAGTTTCATGAAAGAACTTCTTACTACCGCGCGCGCGCG 1081
 QY 160 ThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysVal 179
 DB 1082 CAGGCGCTGGGAGAACTCCATCGGCACACACTGTCCTCAACAGTGCTTCGTGAGGTA 1141
 QY 180 LeuArgAspProSerArgProTyrGlyAspAsnTyrTrpMetLeuAsnProAsnSer 199
 DB 1142 CCGCGCCACTAGGAGCG 1198
 QY 200 GluTyrThrPheAlaAspGlyVal-----PheArgArgArgLysArgLeuSer 216
 DB 1199 GACGACGTGTTTCATCG 1257
 QY 217 HisArgAlaProValProAlaProGlyLeuArgProGluAlaProGlyLeuProAla 236
 DB 1258 GGCCAGCTAGCCTTTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1287
 QY 237 AlaProProAlaProAlaProAlaProAlaProArgMetArgSerProAlaArgGln 256
 DB 1288 GCTCA-CCTCCA-----CCGCGCTCACCTTCATGAGCGCGCGCGCGCGCGCGCG 1328
 QY 257 GluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeu 276
 DB 1329 ---CCCTCTACTGCG 1346
 QY 277 ArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrp 296
 DB 1347 -----CGCGCTTCTGTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1379
 QY 297 GlyAla-----AlaProCysProProLeuProAlaPheProAlaLeuLeu 311
 DB 1380 CGAGCACTTTCAGTTTACACGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1433
 QY 312 ProAlaAlaProCys-----ArgAlaLeuLeuProLeu-----322
 DB 1434 CCTACAGCTCGGTGTTGACTCAAAACTCGCTGGGCGCAACACCACTCTCTCTCCACGCGCA 1493

QY 323 -----CysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla 335
Db 1494 ACGGGCTGAGCGTGGACGGCTGGTCAACGGGGAGATCCGATACGCCACGACCACTCA 1553
QY 336 ArgGluAlaGluValProProThrAlaProProLeuLeuAlaProLeuProAlaAla 355
Db 1554 CGGCCCTCGCTCGCGCTCCGCTCGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1613
QY 356 AlaProAlaAlaProLeuAlaArgGlyProAlaAlaGlyAlaHisLeuTyrCysProLe 375
Db 1614 ACTCCCTCAACCCCTGCTCGCTCAACCTGCTCGCGGGCCAGACCACTTACTTTTCCGCC 1673
QY 375 uArgLeuProAlaAlaLeu-----GlnAlaAlaLeuValArgArgProGlyPr 391
Db 1674 ACCT-----CCGACCCCGTCAATGACTTCGACAGACGACGCTCCATGAGCGCGCGCGCG 1730
QY 391 ohisLeuSerTyr 395
Db 1731 CGTCTCTCTCTAC 1743

RESULT 13

US-08-331-644-4
; Sequence 4, Application US/08331644
; Patent No. 5976872
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,644
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-331-644-4

Alignment Scores:
Pred. No.: 2,72e-11 Length: 1155
Score: 360.50 Matches: 116
Percent Similarity: 45.27% Conservative: 37
Best Local Similarity: 34.32% Mismatches: 118
Query Match: 16.98% Indels: 67
DB: 2 Gaps: 13

US-10-087-080-32 (1-402) x US-08-331-644-4 (1-1155)
QY 26 SerAspAlaProSerPro----LeuSerAlaAlaGlyAspAspSerLeuGlySerAspGly 44
Db 207 TCCTGCTCTCTCGCGGCTCGCTCGCGGCGCGCGAGTCGTCGTCGCGCGCGC 266
QY 45 AspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGlnGlyAspGlyGlu 64
Db 267 GACCGGCGCTCTGGACACAGCATGCGAGGCGG-----TTACTCGGTATCGGACCC 317
QY 65 GlnSerAlaGlyGly-ProGlyAlaGluGluAlaIleProAlaAlaAlaAlaAla 84
Db 318 CAACACCTCGGAGTGTACCTATTATGAGTGAAGAACTACTACCGGCGCGCGCAG 377
QY 84 aValValAlaGluGlyAlaGluAlaGlyProGlyAlaGlyGlyAla----- 102
Db 378 CTAC-----GGCGCATGCGCCACCCCATGGGCGTCTACTCCGGCCACCGGAGCA 428
QY 103 ----GlySerGlyGluGlyAlaArgSerLysProTyrThrArgPro----- 117
Db 429 GTACGGCGCGCATGGCGCTCTACGGCGCCCTACCATCATCAGCCCTTTCTCCAA 488
QY 118 -----LysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSe 134
Db 489 GGACCTGCTGAAGCGCCCTACAGCTATATAGCGCTCATCACCATGCGATCCGAACGC 548
QY 134 rAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPh 154
Db 549 GCCAGAGAAGAGATCACTCTGAACGGCATCTACCACTTCATCATGACCGTTCCCTT 608
QY 154 ePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAs 174
Db 609 CTACCGGAGAACAAAGCAGGCGCTGGCAGAACAGCATCCGCCACCACTGTCATCAATGA 668
QY 174 pCysPheValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTpmE 194
Db 669 GTGCTTGTGAAGTGGCGCGGACGACAAAGACCG---GGCAGGGCAGCTACTGTGAC 725
QY 194 tLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysAr 214
Db 726 GCTCGACCGGACTCTTACACACATGTTTCGAGAATGGCAGCTTCTCGCGGCGCGCGCG 785
QY 214 gLeuSerHisArgAlaProAlaProAlaProGlyLeuArgProGluGlu-AlaProGlyL 234
Db 786 CTTCAAG---AAGAAGATGTGCCCAAGGAC-----AAGAGAGCGCGGCCACCT 833
QY 234 euProAlaAlaProProAlaProAlaAla----- 244
Db 834 CAGGAGCGCGCTCGACACCGGCGGCGCTCCGACAGGACCGCGGTAGCTGACGG 893
QY 245 --ProAlaSerProArgMetArgSerProAlaArgGlnGlu-----A 259
Db 894 GCCCAAGGAGCGCGAGAAAGTCTGTTAAGAGCGAGCGCGCGCTCCCGGCGCTGCC 953
QY 259 rGAlaSerProAlaGlyLysPheSerSerPheAlaIleAsePserIleLeuArgLysP 279
Db 954 GGTATCACCA----- 964
QY 279 roPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTyrGlyAla 299
Db 965 -----AGGTGGAGACGCTGAGCCCGGAGCGCTGAGCGCGCTGCGCG 1010
QY 299 laProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAla 319
Db 1011 CAGGCGATCTCCACGCGCGCGAGTGTCC-----CAGACCGCTCGCTCGCGGAGCA 1061
QY 319 euLeuProLeu-----CysAlaTyrGlyAlaGlyGluProAla 331
Db 1062 CCACGCGCGCGGCTTAACGGGCTCGCGGCTTCAAGCGTGGAGACCATCA 1111

RESULT 14
PCT-US93-04102-4

; Sequence 4, Application PC/TUS9304102

GENERAL INFORMATION:
 APPLICANT: Tac, Wufan
 TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
 ACIDS ENCODING SAME AND USES THEREOF
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: John P. White
 STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/04102
 FILING DATE: 19930430
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/882,292
 FILING DATE: 13-MAY-1992

ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 41472A-PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-977-9550
 TELEFAX: 212-664-0525

INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1155 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 PCT-US93-04102-4

Alignment Scores:
 Pred. No.: 2,72e-11
 Score: 360.50
 Percent Similarity: 45.27%
 Best Local Similarity: 34.32%
 Query Match: 16.98%
 DB: 5

US-10-087-080-32 (1-402) x PCT-US93-04102-4 (1-1155)

26 SerAspAlaProSerPro---LeuSerAlaAlaGlyAspSerLeuGlySerAspGly 44
 207 TCCTGCTCTCTGCTCGGCGCTCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
 45 AspCysAlaAlaLeuProSerAlaGlyGlyAlaAArgAspThrGlnGlyAspGlyGlu 64
 267 GACCGGGGCTGGGACACAGATCCAGCGCGG-----TTACTCGGTATCGGACCC 317
 65 GlnSerAlaGlyGlyGly-ProGlyAlaGluGluAlaAlaProAlaAlaAlaAla 84
 318 CAACACCTGGAGTGGTACCTATTGAGTGAGCAAACTACTACCGGGCGCGCGCGCG 377
 84 aValValAlaGluGlyAlaGluAlaAlaGlyProGlyAlaGlyGlyAla----- 102
 378 CTAC-----GGCGGATGGCCAGCCCATGGCGCTCTACTCCGCGCACCGCGGACCA 428
 103 ----GlySerGlyGluGlyAlaAArgSerLysProTyrThrArgArgPro----- 117
 429 GTACGGCGCGCGCATGGGCGGCTCTCTACGGCGCTTACCACCATCAGCCCTTTCTCCCAA 488
 118 -----LysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleAArgAspSe 134

489 GGACCTGCTGAAGCGCGCTACAGCTATATAGCGCTCATCACCATGGCGCATCAGACGC 548
 134 rAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPh 154
 549 GCAGAGAAGAAGATCACTCTGAACGGCATCTACAGTTCATCATGGACCGTTCCCTT 608
 154 ePheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAs 174
 609 CTACCGCGAGAACAGCAGCGGCTGCAGACAGCATCGCCACACCTGTCTACTCAATGA 668
 174 pCysPheValIysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMe 194
 669 GTGCTTCGTGAAGTGCAGCGCGCAGCAGAACAGGCG---GCCAAGGCGCATCTACTGAC 725
 194 tLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArg 214
 726 GCTCGACCGGACTCTTACCAATGTCAGATGGCAGCTTCCTGCGCGCGCGCGCGCG 785
 214 gLeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGlu-AlaProGlyL 234
 786 CTTCAAG---AAGAAGGATGTGCCAAGGAC-----AAGGAGGAGCGGCGCCACCT 833
 234 euProAlaAlaProProAlaProAlaAla----- 244
 834 CAAGGAGCGCGCTCGACACCGCCAGGGCGCTCCGACAGGAGACCCCGGTAGTGACGG 893
 245 --ProAlaSerProArgMetArgSerProAlaArgGlnGlu-----A 259
 894 GCCCAAGGAGCGCGAGAAAGTCTGTAAAGAGCGAGCGCGCGCTCCCCCGCACTGCC 953
 259 rGAlaSerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuA-gLysP 279
 954 GGTCTATCACA----- 964
 279 roPheArgSerArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaA 299
 965 -----AGTGGAGCGCTGAGCCCGCGAGGAGCGCTCGAGCGCGCGCGCGCG 1010
 299 laProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAla 319
 1011 CAGCGCATCTCCACCGCGCGCGAGTGCTCC-----CAGACGCTCGCTCGCGGAGCA 1061
 319 euLeuProLeu-----CysAlaTyrGlyAlaGlyGluProAla 331
 1062 CCACGCGCGCGCGCTTAACGGCTGCGCGCTTACGGTGGAGACCATCA 1111

RESULT 15

US-09-976-594-495
 ; Sequence 495, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Suchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 495
 ; LENGTH: 5080
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 206866.1
 ; NAME/KEY: unsure
 ; LOCATION: 2468
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-976-594-495

Alignment Scores:

Pred. No.: 1.06e-08 Length: 5080
Score: 324.50 Matches: 106
Percent Similarity: 44.53% Conservative: 29
Best Local Similarity: 34.98% Mismatches: 88
Query Match: 15.28% Indels: 81
DB: 4 Gaps: 11

US-10-087-080-32 (1-402) x US-09-976-594-495 (1-5080)

```
QY 109 ArgSerLysProTyrThrArgArgProLysProProTyrSerTyrIleAlaLeuIleAla 128
DB 23 CQTTCAGCGCAGCTACCGCAGCAAGCGCCCTACTCGTACATCTCGCTCATCACC 82
QY 129 MetAlaIleArgAspSer-AlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLe 148
DB 83 ATGGGCATCCAGCGCGCCCGCCAGCAAGATGCTCAGCTGAGCGAGATCTACCAGTGGAT 142
QY 148 uMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArgHi 168
DB 143 CATGGACCTCTCCCTTATACCGGCAGAACCCAGCAGCGCTGGCAGAACTCCATCCGCCA 202
QY 168 sAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProSerArgProTyrGl 188
DB 203 CTCGCTGTCTCAATGACTGCTTCGTCGAAGTGGCAGCGCTCCCGGACAAGCCG---GG 259
QY 188 VlysAspAsnTyrTyrMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPh 208
DB 260 CAAGGCTCTCTACTGACCGCTGCACCGGACTCCGGCAACATGTTCCGAGACGGCTGCTA 319
QY 208 eArgArgArgArgLysArgLysSerHisArgAlaProValProAlaProGlyLeuArgPr 228
DB 320 CTTGCGCCGC-----CCTCCATCGGGGTGTGCACG 349
QY 228 oGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaSerPr 248
DB 350 GGAAGACCGCCAGCTAG-AGGCGCGCGCGCGCCCGCGCCG-----CCGCCAGCCCC 402
QY 248 oArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSe 268
DB 403 CAGACTCTGGACCACACAGTGGGGCGGCGGACGAGGGGGCGCT----- 445
QY 268 rSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArgArgLeuArgAspTh 288
DB 446 -----CGAGTTGAGACTCCAGCCTCC 468
QY 288 r-----AlaProGlyThrThrLeuGlnTyrGlyAlaAlaPro---Cy 301
DB 469 TCAACTGCGCGCCCATAGCTCCGGCGCG-----GGCGCTGGCTCTGTG 516
QY 301 sProProLeu-ProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuP 321
DB 517 CCGCGCTCTCACCCTGC----- 533
QY 321 roLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAla---ArgGluAlaGluV 340
DB 534 -----ACAGGCTTGGCACCCCGCAGTCCCGAGTCCAGCTG 564
QY 340 alProProThrAlaProProLeuLeuAlaPro----- 351
DB 565 CACCTGAAGGGGACCCCACTACTCTTCAACACCCCGTTCTCCATCAACACCTCATG 624
QY 352 --LeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyAlaHisL 371
DB 625 TCCTCTCGGAGCAGCAGCAGTACGCTGAGCTTCAAGCATACGACAGGCTGCAATAC 684
QY 371 euTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyP 391
DB 685 TC-----GCCTTAGGCTCTACGTTGCCCGCCAGCGCTGCTCTAGGCAGCGCTCGGTG 738
QY 391 roHis 392
|||
```

DB 739 ACCAC 743

Search completed: March 31, 2004, 08:40:24
Job time : 130 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 08:01:11 ; Search time 396 Seconds
(without alignments)
3779.419 Million cell updates/sec

Title: US-10-087-080-32

Perfect score: 2123

Sequence: 1 MKLEVFVRAAHGDKQSDL.....AALVRRPGPHLSYPVETILA 402

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10087080/runat_29032004_114948_24643/app_query.fasta_1.583
-DB=Published Applications_NA_QPWT-fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -ENB=1 -YATRX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10087080 @CGN 1 1 107 @runat_29032004_114948_24643
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOF=6 -FGAPEXT=7 -YGAPOF=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	2123	100.0	1209	12	US-10-229-345-9	Sequence 9, Appli
2	2123	100.0	1209	12	US-10-274-177-9	Sequence 9, Appli
3	2123	100.0	1209	15	US-10-087-080-31	Sequence 31, Appl
4	716	33.7	585	15	US-10-027-632-231353	Sequence 231353
5	488.5	23.0	2487	14	US-10-205-823-135	Sequence 135, App
6	487.5	23.0	2271	14	US-10-101-510-133	Sequence 133, App
7	470	22.1	320	9	US-09-833-381-1266	Sequence 1266, Ap
8	466.5	22.0	2187	10	US-09-960-706-947	Sequence 947, App
9	423	19.9	1662	10	US-09-292-862-1	Sequence 1, Appli
10	419.5	19.8	2753	14	US-10-007-280A-115	Sequence 115, App
11	411	19.4	2593	14	US-10-293-582-24	Sequence 24, Appl
12	411	19.4	3098	14	US-10-198-846-10858	Sequence 10858, A
13	407.5	19.2	1137	14	US-10-029-386-24788	Sequence 24788, A
14	406	19.1	2872	9	US-09-969-708-455	Sequence 455, App
15	406	19.1	2872	14	US-10-177-293-211	Sequence 211, App
16	405	19.1	3289	9	US-09-963-285-8	Sequence 8, Appli
17	405	19.1	6458	9	US-09-963-285-1	Sequence 1, Appli
18	403	19.0	2712	9	US-09-963-285-6	Sequence 6, Appli
19	403	19.0	6021	9	US-09-963-285-5	Sequence 5, Appli
20	391.5	18.4	1506	9	US-09-963-285-9	Sequence 9, Appli
21	391	18.4	4454	9	US-09-764-887-496	Sequence 496, App
22	391	18.4	4454	14	US-10-073-961-496	Sequence 497, App
23	387	18.2	4450	14	US-10-073-961-497	Sequence 497, App
24	387	18.2	4450	14	US-10-029-386-20688	Sequence 20688, A
25	358.5	16.9	7114	14	US-10-388-934-44	Sequence 44, Appl
26	349	16.4	2218	15	US-10-029-386-10008	Sequence 10008, A
27	346	16.3	599	14	US-09-867-701-10873	Sequence 10873, A
28	327	15.4	3119	9	US-09-918-995-4735	Sequence 4735, Ap
29	325.5	15.3	477	10	US-10-027-632-138700	Sequence 138700,
30	314	14.8	425	15	US-10-027-632-138701	Sequence 138701,
31	314	14.8	425	15	US-09-963-285-3	Sequence 3, Appli
32	313.5	14.8	4158	9	US-10-027-632-138703	Sequence 138703,
33	313	14.7	425	15	US-10-027-632-138699	Sequence 138699,
34	310	14.6	425	15	US-10-044-442-10	Sequence 10, Appl
35	309.5	14.6	1605	13	US-10-044-442-13	Sequence 448, App
36	308	14.5	730	9	US-10-044-442-13	Sequence 13, Appl
37	307	14.5	1634	13	US-10-044-442-16	Sequence 16, Appl
38	297	14.0	1668	13	US-10-029-386-23708	Sequence 23708, A
39	295.5	13.9	312	14	US-10-101-510-419	Sequence 419, App
40	277	13.0	1056	14	US-09-960-352-6254	Sequence 6254, Ap
41	273	12.9	441	9	US-10-007-926A-408	Sequence 408, App
42	264	12.4	3059	14	US-10-007-926A-134	Sequence 134, App
43	263	12.4	5723	14	US-10-341-434-52	Sequence 52, Appl
44	263	12.4	5723	15	US-10-062-674-2145	Sequence 2145, Ap
45	260	12.2	3029	15		

ALIGNMENTS

RESULT 1

US-10-229-345-9
; Sequence 9, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-229-345-9

Alignment Scores: 3.79e-142 Length: 1209
Pred. No.: 1209

QY 101 GlyAlaGlySerGlyGluGluValAlaArgSerLysProTyrThrArgArgProLysProPro 120
Db 301 GCGCGGGGAGCGCGAGGGTGCACGACGAAGCATATACGCGCGCGCCCAAGCCCCC 360
QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
Db 361 TACTCGTACATCGCGCTCATCGCATGGCCATCCGCACTCGGGGGCGGGCGCTTACG 420
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
Db 421 CTGGCGGAGATCAACGAGTACCTATGGCAAGTTCCCTTTTCCGCGGCGAGCTACAG 480
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAspCysPheValIysValLeu 180
Db 481 GGCTGGCGCACTCGCTGGCGCACAACTTTCGCTCAACGACTGCTTTCGTCAGGGTGTG 540
QY 181 ArgAspProCysArgProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200
Db 541 CCGGACCCCTCGCGCCCTCGGGCCAGGACAACTACTGGATGCTCAACCCCAACAGCGAG 600
QY 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLeuSerHisArgAlaPro 220
Db 601 TACACCTTCGCGAGCGGGTCTTTCGCGCGCCGCGCAAGCGCTCAGCCACCGCGCGCG 660
QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240
Db 661 GTCCCGCGCCCGGCTGGCGGCCGAGAGGCCCCCGGCGCTCCCGCGCGCCCGCGGCC 720
QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla 260
Db 721 GCGCGCGCGCGCCCGCTCGCGCCGCGCATGCGCTCGCGCGCGCGCGCGAGGAGCGCGCC 780
QY 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
Db 781 AGCCCGCGCGGCAAGTTCCTCCAGCTCTTCGCACTCGACAGCATCTCGCGCAAGCCCTTC 840
QY 281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300
Db 841 CCGACCGCTCGCTCAGGGACACGCGCCCGCGGACGAGCTTCAGTGGGGCGCGCGGCC 900
QY 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeu 320
Db 901 TGCCCGCGCGCTCGCGCGGTTCGCGCGGTCTTCGCGCGCGCGCGCGCGCGCGCGCG 960
QY 321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaAlaArgLeuGlyAlaArgGluAlaGluVal 340
Db 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCTGGGCGCGCGGAGCGCGAGGTG 1020
QY 341 ProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360
Db 1021 CCACCGACCGCGCGCGCGCTCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCG 1080
QY 361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380
Db 1081 CTCGAGAGCG 1140
QY 381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400
Db 1141 CTGACGCGCGCTTAGTCCGNCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
QY 401 LeuAla 402
Db 1201 CTAGCT 1206

RESULT 3

US-10-087-080-31
; Sequence 31, Application US/10087080
; Publication No. US2003023582A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Markowitz, Sanford David
; APPLICANT: Eos Biotechnology, Inc.
; APPLICANT: Case Western Reserve University

; TITLE OF INVENTION: No. US20030235820A1 Methods of Diagnosis of Metastatic Colorect.
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer
; FILE REFERENCE: 018501-000840US
; CURRENT APPLICATION NUMBER: US/10/087,080
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/272,206
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: US 60/281,149
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: US 60/284,555
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: winged helix/forkhead transcription factor (HPH1)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1161)
; OTHER INFORMATION: n = g, a, c or t
; US-10-087-080-31

Alignment Scores:

Pred. No.: 3 79e-142 Length: 1209
Score: 2123.00 Matches: 402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-087-080-32 (1-402) X US-10-087-080-31 (1-1209)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db 1 ATGAAGTTGAGGTTGTTCGTCCTCGCGCGCGCCCGCGGACCAAGCGGACGTGACCTG 60
QY 21 GluGlyAlaGlyGlySerAspAlaProCysProLeuSerAlaAlaGlyAspAspSerLeu 40
Db 61 GAGGGCGCGGGCGGACGACGCGCGCTCCCGCTGTCGCGCGCGGAGACGACTCCCTG 120
QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60
Db 121 GGCTCAGATGGGACTCGCGCGCGCAAGCCCTCCCGCGCGCGCGCGCGCGCGCGCG 180
QY 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluAlaIleProAlaAla 80
Db 181 GCGACCGCGGACAGAGTGGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 81 AlaAlaAlaValValAlaGluGlyAlaGluGlyAlaGluAlaGlyProGlyAlaGly 100
Db 241 GCTCTCAGCGGTGTGGCGGAGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
Db 301 GCGCGGGGAGCGCGAGGTGACCGCGCAAGCCATATACGCGCGCGCGCGCGCGCGCG 360
QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
Db 361 TACTCGTACATCGCGCTCATCGCATGGCCATCGCGACTCGCGCGCGCGCGCGCGCTG 420
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
Db 421 CTGGCGGAGATCAACGAGTACCTCATGGCAAGTTCCCTTTTCCGCGGCGAGCTACAG 480
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeu 180
Db 481 GGCTGGCGCACTCGCTGGCGCGCAACCTTTCGCTCAACGACTGCTTTCGTCAGGGTGTG 540
QY 181 ArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200

Db 541 CGGAGCCCTCGGGGCTGGGCAAGCAACTACTGATGCTCAACCCCAACAGCGAG 600
Qy 201 TTTTPhaAlaaspGlyValPheArgArgArgArgArgArgArgArgArgArgArg 220
Db 601 TACACCTTCGCGCGGGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Qy 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaProProPro 240
Db 661 GTCCCGCGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Qy 241 AlaProAlaProAlaProAlaProAlaProAlaProAlaProAlaProAlaPro 260
Db 721 CG 780
Qy 261 SerProAlaGlyLysPheSerSerSerPheAlaAlaAspSerIleLeuArgLysProPhe 280
Db 781 AGCCCGCGCGCGAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTGCGCAAGCCCTTC 840
Qy 281 ArgSerArgArgLeuArgArgArgArgArgArgArgArgArgArgArgArgArgArg 300
Db 841 CGCAGCGCTGCGCTTCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Qy 301 CysProProLeuProAlaPheProAlaLeuLeuProAlaProCysArgAlaLeuLeu 320
Db 901 TGCCCGCGCGTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Qy 321 ProLeuCysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluVal 340
Db 961 CCGCTCTGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Qy 341 ProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360
Db 1021 CCACCG 1080
Qy 361 LeuArgGlyProAlaAlaGlyAlaHisLeuTyrcysProLeuArgLeuProAlaAla 380
Db 1081 CTCCGAGCG 1140
Qy 381 LeuGlnAlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThrLeu 400
Db 1141 CTGAGCGCGCGCTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
Qy 401 LeuAla 402
Db 1201 CTAGCT 1206

RESULT 4
US-10-027-632-231353/c
; Sequence 231353, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 231353
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-231353
Alignment Scores:
Pred. No.: 1,2e-42 Length: 585
Matches: 136
Conservative: 0
Percent Similarity: 97.84%
Best Local Similarity: 97.84%
Indels: 0
Query Match: 33.73%
Gaps: 0
DB: 15
US-10-087-080-32 (1-402) x US-10-027-632-231353 (1-585)
Qy 264 GlyLysPheSerSerPheAlaAlaAspSerIleLeuArgLysProPheArgSerArg 283
Db 583 GGCAGAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTGCGCAAGCCCTTCGCGAGCGCG 524
Qy 284 ArgLeuArgArgPheAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProPro 303
Db 523 CGCTCAGGAGACAGCG 464
Qy 304 LeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCys 323
Db 463 CTGCGCGCGTTCGCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 404
Qy 324 AlaTyrcysAlaTyGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProThr 343
Db 403 GGTACG 344
Qy 344 AlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGly 363
Db 343 GCGCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
Qy 364 ProAlaAlaGlyGlyAlaHisLeuTyrcysProLeuArgLeuProAlaAlaLeuGlnAla 383
Db 283 CCG 224
Qy 384 AlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThrLeuLeuAla 402
Db 223 GCCTCAGTCCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 167
RESULT 5
US-10-205-823-135
; Sequence 135, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Ganavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Womsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25

256 nGluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLe 276
1239 GCGGCTACGCGCTGCGAGCTGCC-----GCCCTTACGCGCGCGCTCGGCTCTTCG 1289
276 uAAGlySerProPheArgSerArgArg-----LeuArgAspThrAlaPr 290
1290 CCGCGGAGCGCGCGCGCGCGCGCTTCACCGCGACTCGCGCGCGCGCGCGCG 1349
290 oGlyThrLeuGlnTrpGlyAlaAlaProCysProPro-----LeuProAlaPh 307
1350 GCGCACGCGACGCGC-----GGCGCGGAGCTGCGCGGAGCGCGCTTCGCGTACGCGCGC 1406
307 eProAlaLeuLeuPro-----AlaAlaProCysArgAlaLeuLeuProLeuCysAlaTy 325
1407 ACCGCTCG-GCGCGCGCTTACCGCGCGCGCTCGCGCGCTCGCGCGCGCGCG-- 1463
325 rGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAla----- 338
1464 -----GCCCGCGCGCTCAGCGCTGCGCGCGCTCGCGCTTCTCCATCGAGCATCATCGGG 1519
339 -----GluValProProThrAlaProProLeuLeuAl 350
1520 GCGAGCTTGGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTCAG 1579
350 aProLeuProAlaAlaAlaPro-----AlaLysProLeuArgGlyProAlaAlaGlyG 368
1580 GCCTCG-CCCTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 1638
368 yAlaHisLeuTyrcysProLeuArgLeuProAlaAlaLeuGlnAlaLeu-ValArgA 388
1639 C-----TGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1668
388 rgPro---GlyProHisLeuSerTyrcPro 396
1669 GCGCGCGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1697
RESULT 6
US-10-101-510-133
; Sequence 133, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10101.510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 133
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-133
Alignment Scores:
Pred. No.: 6,45e-26 Length: 2271
Score: 487.50 Matches: 177
Percent Similarity: 44.01% Conservative: 36
Best Local Similarity: 36.57% Mismatches: 137
Query Match: 22.96% Indels: 139
DB: 14 Gaps: 21
US-10-087-080-32 (1-402) x US-10-101-510-133 (1-2271)
QY 28 AlaProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAspGly----- 44
Db 64 GCGCGCGCGCGCGG-CCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 122
QY 45 -----AspCysAlaAlaLysPro----- 50

PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 135
LENGTH: 2487
TYPE: DNA
ORGANISM: Homo sapiens
US-10-205-823-135
Alignment Scores:
Pred. No.: 5.94e-26 Length: 2487
Score: 488.50 Matches: 170
Percent Similarity: 47.56% Conservative: 35
Best Local Similarity: 39.44% Mismatches: 147
Query Match: 23.01% Indels: 83
DB: 14 Gaps: 20
US-10-087-080-32 (1-402) x US-10-205-823-135 (1-2487)
QY 14 AspLysGlnLysSerAspLeuGluGlyAlaGlyGlySerAspAlaProSerProLeuSer 33
Db 506 GAAGAGGAGGACGACGAGGCGCGCTGCGCGCGC-CCGCGTGGCTGTCCCGCGCA 564
QY 34 AlaAlaGlyAspAspSerLeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGly 53
Db 565 GCGCGCGCGCGCGCGCGCTGCTGACCGCGGAGGAGCGAGCTGGAGATCTGGAGGAGA 624
QY 54 GlyGlyAlaArg-AspThrGlnGlyAspGlyGlnSerAlaGlyGlyGlyProGlyAl 73
Db 625 GGAGGAGGAGGATGACATCTG-----CTGGCGCGCGCTGCTGGGGGCTCCCGCGCGC 678
QY 73 aGluGluAlaIleProAlaAlaAlaAlaAlaAlaValAlaGluGlyAlaGluAlaG 93
Db 679 CCG 720
QY 93 yAlaAlaGlyProGlyAlaGlyAlaGlySer---GlyGluGlyAlaArgSerTyrcPr 112
Db 721 CG 777
QY 112 oTyrcThrArgArgProGlyProProTyrcSerTyrcIleAlaLeuIleAlaMetAlaIleAr 132
Db 778 GCTGGTG-----AAGCGCGCTACTGCTATATCGCGCTCATCATGATGCGCATCT 828
QY 132 gAspSerAlaGlyGlyArgGluThrLeuAlaGluIleAsnGluTyrcLeuMetGlyLysPh 152
Db 829 GCAGAGCCCAAGACCGCGCTGACGCTGACGAGATCTGTGATTCATCAGCGCGCTT 888
QY 152 eProPhePheArgGlySerTyrcThrGlyTrpArgAsnSerValArgHisAsnLeuSerIe 172
Db 889 CCCTACTACCGGAGAGTTCCTCCGCTGGCAGACAGCATCCGCGCACAACTCTCGCT 948
QY 172 uAsnAspCysPheValLysValLeuArgAspProSerArgProTrpGlyLysAspAsmTy 192
Db 949 CAACGACTGCTTCTGCTCAAGATCCCGCGGAGCGCGCGCGCGCGCGCGCGCG 1005
QY 192 rTrpMetLeuAsnProAsnSerGluTyrcThrPheAlaAspGlyValPheArgArgAr 212
Db 1006 CTGGACCTGACCGCGGAGTCCCGCAGCATGTTTCGACACCGCGAGCTTCTGCGCGGAG 1065
QY 212 gLysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGluAla-- 231
Db 1066 GAAGCGCTTCAAG---CGGAGCGCGCTGCTCCACCGCGCGCGCGCGCGCGCG 1122
QY 232 -----ProGlyLeuProAlaAla----- 237
Db 1123 GCTGCTGCT 1182
QY 238 ----ProProProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArg 256
Db 1183 CTTCCCGCGCGC-GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1238

Db 242 CTTTCATCATCTACAGCATCTGCNCAATCCCTTCGGCATCCGCCGCTCANGGACACGG 301
Qy 289 laProGlyThrThr 293
Db 302 CCCCCGGGAGCACC 315

RESULT 8

US-09-960-706-947
; Sequence 947, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-0105
; CURRENT APPLICATION NUMBER: US/09/960,706
; PRIOR FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 947
; LENGTH: 2187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 U13220
US-09-960-706-947

Alignment Scores:

Pred. No.: 1,936-24 Length: 2187
Score: 456.50 Matches: 157
Percent Similarity: 45.35% Conservative: 48
Best Local Similarity: 34.73% Mismatches: 159
Query Match: 21.97% Indels: 90
DB: 10 Gaps: 18

US-10-087-080-32 (1-402) x US-09-960-706-947 (1-2187)

Qy 8 ProArgAlaAlaHisGlyAspLysGlnGlySerAspLeuGluGlyAlaGlyGlySerAsp 27
Db 79 CGCCCTCCGCGCCGCGGCTCCCTCCGCGGTCCTCCAGATGACACCGAGCGGGGG---CCG 135
Qy 28 AlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySer---AspGlyAspCys 46
Db 136 CGCCGCGCCCGCTCCGCGCGCGGTGCAGCCCGGTCCCGCGGCGCTCCAGGCGCCCTG 195
Qy 47 AlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGlyGlnSer 66
Db 196 ATGAGCCCG 255
Qy 67 AlaGlyGlyGlyProGlyAlaGluGluAlaProAlaAlaAlaAlaAlaValVal 86
Db 256 TCTCTGCTGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 315
Qy 87 AlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGlyGlyAlaGlyGlyGlyGly 106
Db 316 GCGCCCTCGCTCCCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 375
Qy 107 GlyAlaArgSerLysProThrArgArgPro---LysProProThrSerTyrIleAla 125
Db 376 GCCAAGAGCGGAGCTCGGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435
Qy 126 LeuIleAlaMetAlaIleArgAspSerAlaGlyArgLeuThrLeuAlaGluIleAla 145
Db 436 CTATGCTGTCATGGCCATCCAGAGCTCGCCAGCAAGCGCTCAGCTCAGCGAGATCTAC 495
Qy 146 GluTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSer 165
Db 496 CAGTCTCTGAGCGCGCGCTTCCCTCTTCTCCCGCGCGCGCTCCAGCGGCTGGAAGAACTCG 555

Qy 166 ValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProSerArg 185
Db 556 GTGCGCCACAATCTCTCGCTCAACAGAGTCTTCATCAAGCTGCTTAAGGCGCTCGGGCGG 615
Qy 186 ProTyrGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAsp 205
Db 616 CCC---GGCAAGGGCCACTACTGACCATCGACCGCGCGCGCGCGCTTCATGTTTCGAGGAG 672
Qy 206 GlyValPheArgArgArgLysArgLeuSerHisArgAlaPro----- 220
Db 673 GGCTCGTTCGCGCGCGCGCC-GCGCGCGCTTCAGCGCGGAAGTGCACGCGCTCAAGCCCAT 731
Qy 221 ValPro----- 227
Db 732 GTACCAACCGCGTGTGAGCGGCTTGGCTTCGGGGCGTCTGCTGCTCCCGCAGGGCTTCGA 791
Qy 228 ProGluGluAlaProGlyLeuProAlaAlaProProProAlaProAlaProAlaSer 247
Db 792 -----CTTCCAGCGCGCGCGCGCTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCA 833
Qy 248 ProArgMetArgSerPro-----AlaArgGln 256
Db 834 GGGCGGCTACGGCGGCTTCGACATGATGCCCGGGGTAGCAGCGCGCGCGCGCGCGCGCG 893
Qy 257 GluGluArgAla-----SerProAlaGlyLysPheSerSer 268
Db 894 CAGCCAGCGCACCTCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCC 953
Qy 269 SerPheAlaIleAspSerIleLeu-----ArgLysProPhe 280
Db 954 GGGTTCACCTACATGCGCGAGCTGCGCGGTGCGCGGGGACCGCGGGGCGCTCGGTGCGGC 1013
Qy 281 ArgSerArgArgLeuArg-AspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPr 300
Db 1014 CGGGGGCG 1061
Qy 300 CysProProLeuProAlaPhe-----ProAlaLeuLeuPr 312
Db 1062 GGTACCTCTGCTCCCGCGCATGCGCGCGCATCGAATGCCACTGCGCGCTTACGAGCGCC 1121
Qy 312 oAlaAla-----ProCysArgAlaLeuLeuPro-----LeuCysAlaTy 325
Db 1122 TGGGCGCATCTGAGCTGCGCTGG-CGCTCGCTTACTCTCAAGCAGCGCGCTGCGCTGA 1180
Qy 325 rGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThrAlaPr 345
Db 1181 CGCCAGCAGCAACCGCGCGCTCGCGCGG-----CCTGCACTCCAGCA 1225
Qy 345 oProLeuLeuAlaProLeuProAlaAlaAlaProAlaAlaProAlaLysProLeuArgGlyProAl 365
Db 1226 TGCTCTCTACTCGTGGAGCAGAGCTACTTTCACCAAGACGCTCG---CGAGGACCTCT 1282
Qy 365 aAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLe 385
Db 1283 CAGTGAGCATGCC-----CGTTACCAAGCATCACTCTACTCTCAATG 1324
Qy 385 u---ValArgArgProGlyProHisLeuSerTyr 395
Db 1325 GTGACAGAAAAGATTTCGCTCTCACTCAATG 1358

RESULT 9

US-09-232-862-1
; Sequence 1, Application US/09292862
; Publication No. US20030013087A1
; GENERAL INFORMATION:
; APPLICANT: Walter, Tim
; APPLICANT: Jordan, Tim
; APPLICANT: Raymond, Vincent
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE FREAC3 GENE FOR
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT
; TITLE OF INVENTION: DYSENESIS
; FILE REFERENCE: 07540/020003

```

, CURRENT APPLICATION NUMBER: US/09/292,862
, CURRENT FILING DATE: 1999-04-16
, EARLIER APPLICATION NUMBER: 60/084,784
, EARLIER FILING DATE: 1998-05-08
, EARLIER APPLICATION NUMBER: 60/082,206
, EARLIER FILING DATE: 1998-04-17
, NUMBER OF SEQ ID NOS: 6
, SOFTWARE: Fast-SEQ for Windows Version 3.0
, SEQ ID NO 1
, LENGTH: 1662
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-292-862-1

```

Alignment Scores:	1.89e-21	Length:	1662
Pred. No.:	423.00	Matches:	139
Score:	40.13%	Conservative:	42
Percent Similarity:	30.62%	Mismatches:	111
Best Local Similarity:	19.92%	Indels:	159
Query Match:	10	Gaps:	19
DR:			

US-10-087-080-32 (1-402) X US-09-292-862-1 (1-1662)

QY	72	GIYAlaGluGluAlaIleProAlaAlaAlaAlaValValAlaGluGlyAlaGlu	91
DB	58	GGCGGCGAGCAGAGTACTACCGCGCGCGCGCGCG-----GCCGGGGCGCGGTAC	111
QY	92	AlaGlyAlaAlaGlyProGlyAla-----	99
DB	112	ACGCCCATCGCGCCCATGAGCGGTGTCGTCACCTCGGCACGCCGAGCAGTACC	171
QY	100	GlyGlyAlaGlySerGlyGlyGlyAlaArgSerLysProThrArgArgPro-----	117
DB	172	GGCGGCGATCGCGCGCGCTACGGG-----CCCTACACGCGCAGCGCGCAGGCC	219
QY	118	-----LysProProThrSerTyrlleAlaIleAlaMetAlaIleArgAsp	133
DB	220	AAGGACATGGTGAAGCCCGCTATAGCTACATCGGCTCATACCATGGCCATTCAGAAC	279
QY	134	SerAlaGlyGlyArgIleThrLeuAlaGluIleAsnGlyTrpLeuMetGlyLysPhePro	153
DB	280	GCCTCGGACAAGAAGATCACCCTGAACGGCATCTACCACTTCATCATGACCGCTCCCC	339
QY	154	PhePheArgGlySerTyThrGlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsn	173
DB	340	TTCTACCGGACAACAAGCAGCGCTGGCAGAACAGCATCCGCCACACCTCTCGCTCAAC	399
QY	174	AspCysPheValIysValLeuArgAspProSerArgProTrpGlyLysAspAsnTyTrp	193
DB	400	GAGTGCTTCGTCAAGTGCGCGCGACGACACAAGAAGCG--GGCAAGGCGCAGCTACTGG	456
QY	194	MetLeuAsnProAsnSerGlyTrpThrPheAlaAspGlyValPheArgArgArgGlyLys	213
DB	457	ACGCTGGACCGGACTCTTACAACATGTTGCAGAACGCGAGCTTCTTGCGCGCGCGCGCG	516
QY	214	ArgLeuSerHisArg-----	218
DB	517	CGCTTCAAGAAGAAGCAGCGGTTGAAGACAAAGGAGGAGGAGGAGCTGCACCTCAAG	576
QY	219	-----AlaProValProAlaProGlyLeuArgProGluGluAla	231
DB	577	GAGCGCGCCCGCGCGCGCAGCCCCCGCGCGCGCG-----CCGAGCAGAGCGC	627
QY	232	ProGlyLeuProAlaAlaPro-----	238
DB	628	GACGGCAACGCGCCCGGTCCGCAGCGCGCGCGCTCCAGGACATCAAGACCGAG	687
QY	239	-----ProProAlaProAlaProAlaSer-----	247
DB	688	AACGGTACGTGCCCGCCCGACGCCCTGTCTCCCGCGCGCGCGCTTGGGACGGCG	747
QY	248	-----ProArgMetArgSerPro-----AlaArgGlnGluGlu	258

RESULT 10

US-10-007-280A-115
; Sequence 115, Application US/10007280A
; Publication No. US20030059784A1
: GENERAL INFORMATION:

```

: GENERAL INFORMATION :
: APPLICANT: Sun, Tongming
: APPLICANT: Recipon, Herve
: APPLICANT: Salceda, Susana
: APPLICANT: Chenghua, Jiu
: TITLE OF INVENTION: Compositions and Methods
: FILE REFERENCE: DEX-0257
: CURRENT APPLICATION NUMBER: US/10/007,280A
: CURRENT FILING DATE: 2001-11-07
: PRIOR APPLICATION NUMBER: US 60/246,640
: PRIOR FILING DATE: 2000-11-08
: NUMBER OF SEQ ID NOS: 238
: SOFTWARE: Patent Invention version 3.1

```

```
; SEQ ID NO 115
; LENGTH: 2753
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-007-280A-115
```

Alignment Scores:
Pred. No.: 5,09e-21
Score: 419.50
Percent Similarity: 40.73%
Best Local Similarity: 33.41%
Query Match: 19.76%
DB: 14

Length:	2753
Matches:	155
Conservative:	34
Mismatches:	143
Indels:	132
Gaps:	17

[illegible]

```

RESULT 10
US-10-007-280A-115
; Sequence 115, Application US/10007280A
; Publication No. US20030059784A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Chenghua, Liu
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pri
; FILE REFERENCE: DEX-0257
; CURRENT APPLICATION NUMBER: US/10/007,280A
; CURRENT FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,640
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 2753
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-007-280A-115

Alignment Scores:
Pred. No.: 5,09e-21
Score: 419.50
Percent Similarity: 40.73%
Best Local Similarity: 33.41%
Query Match: 19.76%
Gaps: 14
Matches: 155
Conservative: 34
Mismatch: 143
Indels: 132
Gaps: 17

```

US-10-087-080-32 (1-402) x US-10-007-280A-115 (1-2753)

15 LysGlnGlySerAspLeuGluGlyAlaGlyGly----- 25
84 CGACAGGCTTCAGCAAGCAAGCGCGGCATCCGAGTCTCCAGAAATTGAGACT 143
26 -----SerAspAlaProSerPro-LeuSerAlaAlaGlyAspAspSerLeuGlyse 42
144 TGGCCGTAAGCGAGCTCGTGGCCCAACTCTTTCGCGCGCCAGCGCTCGAGGGAGAG 203
42 rAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAs 62
204 CAGAGCGCGCGCGCGG-----GCGCGCGCGCTTGTGATGA 242
62 pGlyGlu-----GlnSerAlaGlyGlyProGlyAlaGluGluAlaIlePr 78
243 TGGCCAGTACCCCGAGCGCGGCGCGCGG-----CC 281
78 oAlaAlaAlaAlaAlaValVaAlaGluGlyAlaGluAlaAlaAlaGlyProG 98
282 TGTGCGCCCGAGACCGTCTCAGTCAAGAGGCGGAGAGCGCGCGCGAGCCAG 341
98 lYAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProL 118
342 GCAAGCGCGTGGGGTGGCGCGGAGCAGCCCGGAGAGCGGAGCGCGCGAG---A 398
118 ySProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyAla 138
399 AGCCCGCTACTGTAGTGGCGCTCATCGCCATGCGATCGCGAGAGCGCGGAGAAGA 458
138 rGLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlyLys 158
459 GGCTCAGCTGTCGGGATCTACGATCATCGGAGATTCGCTTCTACGAGAAGA 518
158 erTyrThrGlyTyrArgAsnSerValArgHieAsnLeuSerLeuAsnAspCysPheValL 178
519 ATAAGAAGGCTGCGCAAAATAGCATCGCCCAACCTCAGCTCAACGAGTCTCATCA 578
178 ySValLeuArgAspProSerArgProTyrGlyLysAspSerTyrTrpMetLeuAsnProA 198
579 AGGTGCGCGGAGCGCGCGCGAG---CGCAAGGGCACTACTGAGCGCTGGACCGG 635
198 snSerGluTyrThrPheAlaAspGlyValPheArgArgArg-----LysArg-Leu 215
636 CTGCGAGACATGTCGAGAGGGCACTACCGCGCGCGCGCGCATGAAGAGCGCCT 695
216 SerHieArgAlaPro-----ValProAlaProGlyLeuArgProGluGluAlaPro 232
696 TCGCGCGCGCGCGCGCGCACTTCCAGCGCGCAAGGGGCTTTCGGGCGCGAGCGCGG 755
233 GlyLeuProAlaAla----- 237
756 CAGCGGCGTGGCGTGGCGGCGCGCGCGCGCGCGCTACGCTACGCTACGCGCGCGCG 815
238 -----ProProPro 240
816 AGTACCTGCGAGTGGCTTCTCAACACTCGTGGCGCTACCGCAGCGCTCCCTCACCA 875
241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArg---GlnGluGluArg 259
876 TGGCCTATGCTCTGCG-----AGATGGCGGAGCGCGCGCGGTGAGAGCTGCGG 929
260 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 279
930 CTGACGCGCGCGCGCGGTAGCC-----CTGGCGCGCGCG 965
280 PheArgSerArgArgLeuArgAspThrAlaProGlyThr----- 292
966 CTGTGCTCAAGGGCTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025
293 -----Thr-LeuGlnTrpGlyAlaAlaProCysProPr 303

1026 GCATGGCGGTGCGCGCGCGGTAGTAGTAAGTCTGACATGGCTGGGAGCGCGCGCGG 1085
303 oLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeuLeuProLeuCy 323
1086 CACCGCGCGCTCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
323 sAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAla-GluValProPro- 342
1120 -----GCAGCGACCATCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1169
343 -----ThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaProA 358
1170 ACCAGCG 1229
358 lAlysProLeuArgGlyProAlaAlaGlyAlaHisLeuTyrCysProLeuArgLeuP 378
1230 CG 1258
378 roAlaAlaLeuGlnAlaAlaAlaLeuValArgArgProGly-----ProHisLeuSert 395
1259 --CGGCGCTGAGTTCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1316
395 Yr 395
1317 AC 1318

RESULT 11
US-10-293-582-24
; Sequence 24, Application US/10293582
; Publication No. US20030175253A1
; GENERAL INFORMATION:
; APPLICANT: Akil, Huda
; APPLICANT: Bunney, William E.
; APPLICANT: Burke, Sharon
; APPLICANT: Choudary, Prabhakara V.
; APPLICANT: Cox, David R.
; APPLICANT: Evans, Simon
; APPLICANT: Jones, Edward G.
; APPLICANT: Li, Jun
; APPLICANT: Lopez, Juan F.
; APPLICANT: The Trustees of The Leland Stanford Junior University
; TITLE OF INVENTION: Compositions and Methods for Diagnosing and Treating
; TITLE OF INVENTION: Mental Disorders
; FILE REFERENCE: 020885-000210US
; CURRENT APPLICATION NUMBER: US/10/293,582
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/339,252
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 2559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human brain factor 1, HBF-1 transcription factor
US-10-293-582-24

Alignment Scores:
Pred. No.: 1,91e-20 Length: 2559
Score: 411.00 Matches: 131
Percent Similarity: 40.46% Conservative: 28
Best Local Similarity: 33.33% Mismatches: 145
Query Match: 19.36% Indels: 89
DB: 14 Gaps: 12

US-10-087-080-32 (1-402) x US-10-293-582-24 (1-2559)

QY 28 AlaProSerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAspGlyAspCysAla 47
DB 419 GCG 457
QY 48 AlaLysProSer-AlaGlyGlyAlaArgAspThrGlnGlyGluGlnSerAl 67

Qy	184	SerArgProTIPgLyVaAspAsnTyrTirMetLeuAsnProAsnSerGluTyrThrPhe	203
Db	796	GACAAGCGG--GGCAAGGGCTCTACTTGGACGCTGCACCGGACTCCGGCAACATGTC	852
Qy	204	AlaAspGlyValPheArgArgArgLysArgLeuSer---HisArgalaPro-ValPr	222
Db	853	GAGAACGGTGTACTTGCCTCGCGCCAGAAAGCGCTTCAAGTGCAGAAAGCGCGGGGCC	912
Qy	222	oAlaProGlyLeuAsgProGluGluAlaProGlyLeuPro-----Al	236
Db	913	GGCGGGGGGGGAGCGGGAAGCGGGGCAGCGGCCCAAGGGCGGCCCTCAGAGCGCG	972
Qy	236	alaProProProlaProAlaProAlaProAlaSerProArgMetArg-SerProAlaArg	256
Db	973	MAGGACCCCTCTGGGGCGCTCTAAACCCAGCGCGACTCGCCCTCCATCGGGGTGCAC	1032
Qy	256	lnGluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAaspSerIleL	276
Db	1033	GGGAGACCGG-----	1043
Qy	276	euArgLysProPheArgSerArg-----ArgIeuArgAaspThrAlaProGlyThr	293
Db	1044	-----CCAGCTAGAGGGGCGCGCGCCCGCGCGCGCGCCAGCCCCAGACTCTG	1095
Qy	293	hrLeuGlnTIPgLyAlaAlaProCysProProLeuProAla---PheProAlaLeuLeup	312
Db	1096	GACCACAGTGGGGCGACGGCGACAGGGGGCGCTTCGGAGTTGAGACTCCAGCTCTCTCA	1155
Qy	312	roAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyLeuProAla	332
Db	1156	ACTGGCGCCCCATAAG-----CTCCGGGCGCGGGCGGTGGCTCTGTGCC	1203
Qy	332	rgLeuGly-----AlaArgGluAlaGluValProProThrAlap	345
Db	1204	GCCTCTACCCGGCACAGCGTTGGCACCCAGTCCCGAGTGCACCTGAAGGGGAC	1263
Qy	345	rrProLeuLeuAlaProLeuProAlaAlaAlaProAlaLysProLeuArgGlyProAla	365
Db	1264	CCCCACTACTCTTCAACACACCCGTTCTCCATCAACAACCTCATGTCTCTCTCGAGCAG	1323
Qy	365	laAla--GlyGlyAlaHisLeu-----TyrCysProLeuArgLeup	378
Db	1324	CAGCATAGCTGGACTGCATATAACCGAACAGCGCACTGCATATCTCGGCTTACGGCTCT	1383
Qy	378	roAlaAlaLeuGlnAlaAlaLeuValArgArgProGlyProHis	392
Db	1384	ACGTGTCGGCGGACGCTGCTCTAGGCGAGCGCTCGGTGACCAAC	1427

RESULT 13
US-10-029-386-24789/c
; Sequence 24788, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24788
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121790.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2

```

; OTHER INFORMATION: NT HIT: g113631739, EVALUJE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P53117, EVALUJE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: P531970.1, EVALUJE 0.00e+00
US-10-029-386-24788

Alignment Scores:
Pred. No.: 1.64e-20 Length: 1137
Score: 407.50 Matches: 140
Percent Similarity: 44.97% Conservative: 39
Best Local Similarity: 35.18% Mismatches: 152
Query Match: 19.19% Indels: 67
DB: 14 Gaps: 15

US-10-087-080-32 (1-402) x US-10-029-386-24788 (1-1137)
QY 24 GlyGlySerAlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
Db 1079 GGCTCGCGGGCGCCATGAACAGCATGACTGCGCGCGGTGACGCCATGGGTACG---1023
QY 44 GlyAspCysAlaAlaIysProSerAlaGlyGlyGlyAlaAargAspThrGlnGlyAspGly 63
Db 1022 -----GCGGTAGCCCGAGC-----GGCATGGCGGCCATGGGTGCGCAGAGCGGCC 975
QY 64 GluGlnSerAlaGlyGlyProGlyAlaGluGluAlaIleProAlaAlaAlaAla 83
Db 974 TCCATGAAT---GGCCTGGCGCCCTACGCGCGCCCATGAACCGTGCATGACGCCCATG 918
QY 84 AlaValValaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyAlaGly 103
Db 917 GCGTAGCGCGCGTCCAACTTGGCGCGCAGCGCGCGGC-----GGCGCGCGCGAC 867
QY 104 SerGlyGluGlyAlaAargSerLysProTyrThrArgArgProLysProTyrSerTyr 123
Db 866 GCCAAGACGCTTCAAGCGCAGTACCGCGAGCC-----AAGCGCGCTACTCGTAC 816
QY 124 IleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGlu 143
Db 815 ATTCGCTCATCACCATGCCATCCAGAGGGCGCCAGCAAGATGCTCACTGACGCGAG 756
QY 144 IleAsnGlnTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArg 163
Db 755 ATTCACAGCTGGATCATGACCTTCCCTATTACCGCAGACACGACGCGCTGGCAG 696
QY 164 AsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspPro 183
Db 695 AACTCATCGGCACATCGCTGTCTTCAATGATCTGCTCAAGGTGGCAGCGCTCCCG 636
QY 184 SerArgProTropGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlnTyrThrPhe 203
Db 635 GACAAACCG--GGCAAGGGCTCTTACTGACGCTGCACCGGACTCCGGCAACATGTT 579
QY 204 AlaAspGlyValPheArgArgArgArgLysArgLeuSer--HisArgAlaPro-ValPr 222
Db 578 GAGAACCGGTGCTACTTGTGCGCGCCAGAGACGCTTCAAGTGCAGAGACGACGCGGG 519
QY 222 OAlaProGlyLeuArgProGluGluAlaProGlyLeuPro-----Al 236
Db 518 GCGCGCGGGGGAGGAGCGGGGCGCAGCGCGCCAGCGCGCCCTGAGAGCCGC 459
QY 236 alaProProProAlaProAlaAlaProAlaSerProArgMetArg-SerProAlaAarg 256
Db 458 AAGACCCCTCTGGCGCCCTCTAACCCACGCGCGACTCGCCCTCCATCGGGGTGTGCAC 399
QY 256 lngLuuArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleL 276
Db 398 GGGAGACCGG----- 388
QY 276 euArgLysProPheArgSerArg-----ArgLeuArgAspThrAlaProGlyThrT 293
Db 387 -----CCAGCTAGAGGGCGCGCGCCCGCGCGCGCCAGCGCCCGCAGACTCTCG 336
QY 293 hrlLeuGlnTrpGlyAlaAlaIaProCysProProProAlaProAlaPheProAlaLeu 312

```

Db 335 GACCACAGTGGGCGGACGACGAGGGGGCTCGGAGTTCAAGACTCCAGCTCCTCA 276
 Qy 312 roAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyGlyAlaGlyGluProAlaA 332
 Db 275 ACTGCCGCCCCATAAG-----CTCGGGCGGGGGCGCTGCTCTGTGGCC 228
 Qy 332 rGLeuGly-----AlaArgGluAlaGluValProProThrAlaP 345
 Db 227 GCCTCTCACCGGCGACACGAGTGGCACCCACGAGTCCAGCTGCACCTGAAGGGGAC 168
 Qy 345 roProLeuLeuAlaPro-----LeuProAlaAlaA 356
 Db 167 CCCACTACTCTTCAACACCCGTTCTCCATCAACACCTCATGCTCTCTCGAGGAG 108
 Qy 356 laProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysProLeuA 376
 Db 107 CAGCATAAGCTGACCTTCAAGGCATACGACGAGCTGCAATCTC-----GCCATTAC 54
 Qy 376 rGLeuProAlaAlaLeuGlnAlaAlaLeuValArgProGlyProHis 392
 Db 53 GGCTCTAGTGTGCCGCCAGCTGCTCTAGGACGAGCGCTCGGTGACCAC 4

RESULT 14
 US-09-969-708-455
 ; Sequence 455, Application US/09969708
 ; Patent No. US20020102532A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Augustus, Meena
 ; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
 ; FILE OF INVENTION: Sets
 ; FILE REFERENCE: 689290-70
 ; CURRENT APPLICATION NUMBER: US/60/237,606
 ; PRIOR FILING DATE: 2001-10-03
 ; PRIOR APPLICATION NUMBER: US/60/237,606
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: US/60/237,608
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: US/60/237,425
 ; PRIOR FILING DATE: 2000-10-03
 ; NUMBER OF SEQ ID NOS: 658
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 455
 ; LENGTH: 2872
 ; TYPE: DNA
 ; ORGANISM: Homosapiens
 ; US-09-969-708-455

Alignment Scores:
 Pred. No.: 4,79e-20 Length: 2872
 Score: 406.00 Matches: 142
 Percent Similarity: 45.64% Conservative: 41
 Best Local Similarity: 35.41% Mismatches: 146
 Query Match: 19.12% Indels: 75
 DB: 9 Gaps: 13

US-10-087-080-32 (1-402) x US-09-969-708-455 (1-2872)

Qy 24 GlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
 Db 349 GGCTCGGGCGGCCATGACAGCATGACTGGGGCGGCGGTGACGGCCATGGTACG--- 405
 Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGlnGlyAspGly 63
 Db 406 -----GCGGTAGCCCGAGC---GGCATGGGCGCCATGGGTGGCGACGAGCGGCC 453
 Qy 64 GluGlnSerAlaGlyGlyProGlyAlaGluGluAlaAlaLeuValArgProGlyProHis 83
 Db 454 TCCATGATGATGGCTTGGGCGCCCTACGGCGCGCCATGAACCCGTGATGAGCCCATG 513
 Qy 84 AlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyGlyAlaGly 103
 Db 514 GCGTACGCGCGCTCCAACTGGGCGCGACGCGCGGGGC-----GGCGGGCGGAC 564

RESULT 15
 US-10-177-293-211
 ; Sequence 211, Application US/10177293
 ; Publication No. US20030124128A1
 ; GENERAL INFORMATION:

Qy 104 SerGlyGluGlyAlaArgSerLysProTyThrArgArgProLysProProTySerTy 123
 Db 565 GCAAGAGCTTCAAGCGGAGTTACCCGACGCC-----AAGCCGCCCTACTCTGTAC 615
 Qy 124 lleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGlu 143
 Db 616 ATCTGCTCATCACCATGGCCATCGAGCGGGCGGCCAGCAAGATGCTCAAGCTGAGCGAG 675
 Qy 144 lleAsnGluTyLeuMetGlyLysPhePheArgGlySerTyThrGlyTTPArg 163
 Db 676 ATCTACAGTGGATCATGGACCTTCTCCCTATTATTCGGCAGACACAGCAGCGCTGGCAG 735
 Qy 164 AsnSerValArgHisAsnLeuSerLeuAsnAspCysPheVallysValleuArgAspPro 183
 Db 736 AACTCCATCCGCCCATCTCGCTCTCAATGACTGCTTCTCAAGTGGCAGCGCTCCCGG 795
 Qy 184 SerArgProTTPGlyLysAspAsnTyTTPMetLeuAsnProAsnSerGluTyThrPhe 203
 Db 796 GACAGCGG---GGCAGGGGCTCTACTGGACGCTGCACCGGACTCCGCCAACATGTTT 852
 Qy 204 AlaAspGlyValPheArgArgArgGlyArgLeuSerHisArgAlaProValProAla 223
 Db 853 GAGAACGGCTGTCTTGGCGCGCAGAAAGCGC----- 885
 Qy 224 ProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAla 243
 Db 886 -----TTCAAGTGGAGAACGACGCGGGGG---CCGGCGGGGGGGCGGAGCGG 937
 Qy 244 AlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAla 263
 Db 938 GGGGCGAGCGGCGCAAGG-----CGGCGC-CCTGAGAGCGCGCAAGGACCCCTCT 984
 Qy 264 GlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArg 283
 Db 985 GGC-----GCTCTAACCCCGCGCGACTCGCCCTCCATCGGGGTGGCAC----- 1032
 Qy 284 ArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAlaProCysProPro 303
 Db 1033 -----GGGAAGACCGGCGAGCTAGAGGGCGCGCGCGCGCGCGCGCG 1071
 Qy 304 LeuProAlaPheProAlaLeuLeu----- 311
 Db 1072 CCGGCG 1131
 Qy 312 -----ProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyGlyAla 327
 Db 1132 GAGTTGAAGACTCCAGCTCTCTCAACTGC-GCCCCCATTAAGCTCGGGCGCGCGCT 1190
 Qy 328 GlyGluProAlaArgLeuGly-----AlaArgGluAlaGluVal 340
 Db 1191 GGCTCTGTGCGCGCTCTACCGCGGACACGCTTGGCACCACCGAGTCCCGAGTGCA 1250
 Qy 341 ProProThrAlaProProLeuLeuAlaPro----- 351
 Db 1251 CCTGAAGGGGAGCCCCCCTACTCTTCAACACCCCGCTTCCATCAACACTATGTG 1310
 Qy 352 LeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeu 371
 Db 1311 CTCCTCGGAGCAGCAGCATAGCTGGACTTCAAGGCATACGAACAGGCACTGCAATCTC 1370
 Qy 372 TyrCysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgProGlyPro 391
 Db 1371 -----GCTTACGGCTCTAGTTGGCGCGCGCGCTCTAGGACGCGCGCTGAG 1424
 Qy 392 His 392
 Db 1425 CAC 1427

APPLICANT: Lillie, James
APPLICANT: Glat, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhang
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIORITY APPLICATION NUMBER: US 60/299,887
PRIORITY FILING DATE: 2001-06-21
PRIORITY APPLICATION NUMBER: US 60/301,572
PRIORITY FILING DATE: 2001-06-27
PRIORITY APPLICATION NUMBER: US 60/306,501
PRIORITY FILING DATE: 2001-07-18
PRIORITY APPLICATION NUMBER: US 60/325,002
PRIORITY FILING DATE: 2001-09-25
PRIORITY APPLICATION NUMBER: US 60/362,585
PRIORITY FILING DATE: 2002-03-05
PRIORITY APPLICATION NUMBER: US 60/xxx,xxx
PRIORITY FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 211
LENGTH: 2872
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-293-211

Alignment Scores:
Pred. No.: 4,79e-20 Length: 2872
Score: 406.00 Matches: 142
Percent Similarity: 45.64% Conservative: 41
Best Local Similarity: 35.41% Mismatches: 146
Query Match: 19.12% Indels: 75
DB: 14 Gaps: 13

US-10-087-080-32 (1-402) x US-10-177-293-211 (1-2872)

Qy 24 GlySerAspAlaProSerProLeuSerAlaAlaGlyAspSerLeuGlySerAsp 43
Db 349 GGCTCGCGGGCCCATGAACAGCATGACTCGCGCGGGGTGACGGCCATGGGTACG--- 405
Qy 44 GlyAspCysAlaAlaLysProSerAlaGlyGlyAlaAlaArgAspThrGlnGlyAspGly 63
Db 406 -----CGCTGAGCCCGAGC---GGATGGCGCCCATGGTGGCGAGCGGGCC 453
Qy 64 GluGlnSerAlaGlyGlyProGlyAlaGluGluAlaAlaAlaAlaAlaAlaAla 83
Db 454 TCCATGATGAATGGCTGGGCCCTAGCGCGCGCCCATGAACCGGTGATGAGCCCATG 513
Qy 84 AlavalValaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyAlaGly 103
Db 514 GCCTACCGCGCGTCCACCTGGGCGCGAGCGCGCGGGC-----GGCGCGCGCGAC 564
Qy 104 SerGlyGluGlyAlaArgSerLysProTyrThrArgProLysProTyrSerTyr 123
Db 565 GCCAGACGTTCAAGCGCACTTACCCGACGCC-----AAGCGCGCTACTCGTAC 615

Qy 124 IleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThrLeuAlaGlu 143
Db 616 ATCTCGCTCATCACCATTGCCAGCGGGCGCCAGCAAGATGCTCAGCTGAGCGAG 675
Qy 144 IleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThrGlyTyrArg 163
Db 676 ATCTACCATGGATCATGGACCTCTTCCCTATTACCGGACAGAACCCAGCGCTGGCAG 735
Qy 164 AsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspPro 183
Db 736 AACTCCATCGCCACTCGCTGTCTTCAATGACTGTTCTGTTCAAGTGGCAGCGTCCCGC 795
Qy 184 SerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPhe 203
Db 796 GACAAGCCG---GGCAAGGGCTCTTACTGACCTGACCCCGGACTCCGGCAACATGTTT 852
Qy 204 AlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaProValProAla 223
Db 853 GAGAACGGTGTCTACTTTCGCGCGCCAGAGCGC----- 885
Qy 224 ProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAla 243
Db 886 -----TTCAGTTCGAGAAAGCAGCGGGG---CCGCGCGGGGGCGGAGCG 937
Qy 244 AlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAla 263
Db 938 GGGCAGCGGGCCAGG-----GCGGC-CTGAGAGCCGACAGGACCCCTCT 984
Qy 264 GlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArg 283
Db 985 GGC-----GCCTTAACCCAGCGCGCTGCGCTCCATCGGGGTGTGCAC----- 1032
Qy 284 ArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAlaProCysProPro 303
Db 1033 -----GGGAAGCCGGCCAGCTAGAGGCGCGCGCGCGCGCGCG 1071
Qy 304 LeuProAlaPheProAlaLeuLeu----- 311
Db 1072 CCGCGCGCGAGCCCGCAGACTCTGACACACAGTGGGGCGAGCGAGCGGGCGCGCTCG 1131
Qy 312 -----ProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAla 327
Db 1132 GAGTTGAGACTCCAGCTCTCTCAACTGC-GCCCCCATTAAGTCCGGGCCCGGGGGCT 1190
Qy 328 GlyGluProAlaArgLeuGly-----AlaArgGluAlaGluVal 340
Db 1191 GGCCTCTGTGCCCGCTCTCACCGGCACACGGGTGGCAGCCCGCAGTCCAGCTGCA 1250
Qy 341 ProProThrAlaProProLeuLeuAlaPro----- 351
Db 1251 CCTGAAAGGGGAGCCCGCCACTACTCTTCAACCCCGCTTCTCCATCAACACCTCATGTC 1310
Qy 352 LeuProAlaAlaAlaProAlaLysProLeuArgGlyProAlaAlaGlyGlyAlaHisLeu 371
Db 1311 CTCTCGGAGCAGCAGCATAGCTGGACTTCAAGGCATACGACAGGCACTGCAATATCT 1370
Qy 372 TyrCysProLeuArgLeuProAlaAlaLeuGlnAlaLeuValArgArgProGlyPro 391
Db 1371 -----GCCTTACGGCTCTAGCTTGGCGCGCAGCGCTGCTCTAGGCGCGCGCTCGGTGAC 1424
Qy 392 His 392
Db 1425 CAC 1427

Search completed: March 31, 2004, 09:43:26
Job time : 417 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 31, 2004, 05:50:04 / Search time 2579 Seconds
(without alignments)
4654.749 Million cell updates/sec

Title: US-10-087-080-32
Perfect score: 2123
Sequence: 1 MKLEVFVPRAAHGDKQGS...AALVRRPGHLSYPVETLLA 402

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool_p/US10087080/runat_29032004_114946_24430/app_query.fasta_1.583
-DB=EST -OPM=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -TRANS-human40.cdi -LIST=45
-UNITS-bits START=1 -END=1 -MATRIX=blosum62 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080.GCEN_1_13549 @runat_29032004_114946_24430 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	954	44.9	904	13	BQ922461
2	693	32.6	1117	14	CK028187
3	682	32.1	946	14	CD754754
4	676	31.8	799	14	CF997203
5	656.5	30.9	559	9	AI169632
6	615	29.0	804	14	CK030218
7	614	28.9	581	13	BQ285521
8	613.5	28.9	805	13	EX084216
9	611	28.8	564	13	BQ449254
10	586.5	27.6	514	10	BF282916
11	586	27.6	682	9	AL636071
12	566	26.7	878	29	CNS044M2
13	563.5	26.5	398	9	AI586081
14	560.5	26.4	688	12	BI443539
15	519	24.4	542	14	CD282719
16	507.5	23.9	518	10	BB637563
17	503.5	23.7	571	13	BQ615802
18	502	23.6	557	12	BM531626
19	497.5	23.4	590	12	BM573553
20	497.5	23.4	592	12	BM573447
21	496	23.4	596	12	BM586827
22	493	23.2	425	10	BF290883
23	493	23.2	580	12	BM574529
24	493	23.2	582	12	BM530633
25	493	23.2	584	12	BM585052
26	490.5	23.1	556	12	BM573325
27	490.5	23.1	557	12	BM532671
28	490.5	23.1	564	12	BM532182
29	490.5	23.1	585	13	BQ480599
30	490.5	23.0	541	12	BM572533
31	488.5	23.0	560	12	BM574466
32	488.5	23.0	562	12	BM531746
33	488.5	23.0	571	12	BM532199
34	487.5	23.0	556	12	BM572314
35	487.5	23.0	570	12	BM574490
36	487.5	22.9	564	12	BM574918
37	486.5	22.9	570	12	BM532720
38	486.5	22.9	571	12	BM573745
39	486.5	22.9	565	12	BM572374
40	485.5	22.9	570	12	BM532670
41	485.5	22.8	554	12	BM574905
42	484.5	22.8	533	13	BQ615301
43	483.5	22.8	589	12	BM574575
44	483.5	22.8	561	12	BI709264
45	483	22.8			

ALIGNMENTS

RESULT 1
BQ922461
LOCUS
DEFINITION BQ922461 904 bp mRNA linear EST 20-AUG-2002
IMAGE:6200329 5', mRNA sequence.
ACCESSION BQ922461
VERSION BQ922461.1 GI:22337492
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 904)

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LAM13515 row: a column: 02
 High quality sequence stop: 523.
 Location/Qualifiers
 1. 904
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6200329"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_hosts="DH10B"
 /clone_lib="Lupski sciatic nerve"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: Noti; Site 2: SalI; cDNA made by oligo-dr priming. Directionally cloned using the following adaptors: 5'-TCGACCGGCGGCGCG-3' and 5'-GACTAGTCTAGATCGGCGCGGCGGCT(15)-3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

ORIGIN

Alignment Scores:
 Pred. No.: 3.63e-32 Length: 904
 Score: 954.00 Matches: 199
 Percent Similarity: 70.10% Conservatives: 5
 Best Local Similarity: 68.38% Mismatches: 29
 Query Match: 44.94% Indels: 58
 DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ922461 (1-904)

164 AnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeuArgAspPro 183
 |||||
 3 AACTCCGTGCGCCACACCTTTCGCTCAACGACGTCTTGTCAAGGTGCTGCGCGACCCC 62
 |||||
 184 SerArgProTTPGLyLysAspAsnTyrTTPMetLeuAsnProAsnSerGluTyrThrPhe 203
 |||||
 63 TCGCGGCGCTGGGCAAGGACACTACTGGATGCTCAACCCACACGAGTACACCTTC 122
 |||||
 204 AlaAspGlyValPheArgArgArgLysArgLysArgLysSerHisArgAlaProValProAla 223
 |||||
 123 GCCGACGGGGTCTTCCGCGCGCGCGCGAAGCGGCTCAGCCACCGCGCGCGGTCCCCGCG 182
 |||||
 224 ProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAla 243
 |||||
 183 CCCGCGGTGCGCGCGCGAGGCGCCCGCGGCTCTCCCGCGCGCGCGCGCGCGCGCGCC 242
 |||||
 244 AlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAlaSerProAla 263
 |||||
 243 GCCCGCGGCTCGCCCGCATCGCTCGCCCGCGCGCGCGAGGAGCGCGCGCGCGCGCG 302
 |||||
 264 GlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPheArgSerArg 283
 |||||
 303 GGCAGATTCTCCAGTCTCGCCATCGACAGCATCTCGCCAGACGCTTCCGCGAGCGCG 362
 |||||
 284 ArgLeuArgAspThrAlaProGlyThrThrLeuGlnTTPGlyValAlaAlaProCysProPro 303
 |||||

Db 363 CGCCTCAGGGACACGGCCCCCGGGAGCAGCGTTCAGTGGGGCGCGCGCCCTGCCCGCGC 422
 Qy 304 LeuProAlaPheProAlaLeuLeuProAlaAlaProCysA:GAlaLeuLeuProLeuCys 323
 |||||
 Db 423 CTGCCCCGGTTCCTCCCGCGGTCTCTCCCGCGGTGCGCTGAGGGCCCTCTGCGGCTCTGC 482
 Qy 324 AlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProProThr 343
 |||||
 Db 483 GCGTACGGCGGGGAGCGCGCGGTGGCGCGCGCCGCCGCCGCCGCCGCCGCCGCCGCC 542
 Qy 344 -----AlaProProLeuLeuAlaPro 351
 |||||
 Db 543 CCGGCGCGCGCGGTTCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602
 Qy 352 -LeuProAlaAlaAlaProAlaLysProLeuArg-GlyProAlaAlaGlyGlyAlaHisL 371
 |||||
 Db 603 TCTCCCG 659
 Qy 371 euTyr----- 372
 |||||
 Db 660 TGTTCCTCCCGCTCCCGCTATCCCGCGCTTTGTTCCTTTGTTCCTTGAATGCCNC 719
 Qy 372 ----- 372
 Db 720 GGGCCCCGAGGCTCGCGATGGAGGAGCATAGAGCGCGCGCGCGCGCGCGCGCGCGCG 779
 Qy 373 --CysProLeuArgLeuProAlaAlaLeuGlnAlaAlaLeuValArgArgProGly---P 391
 |||||
 Db 780 CCGTGTCCCGCGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 830
 Qy 391 roHisLeuSerTyrProValGluThrLeu 400
 |||||
 Db 831 TTCACCTGTCCCG 859
 RESULT 2
 CK028187 1117 bp mRNA linear EST 26-NOV-2003
 LOCUS AGENCOURT 16624312 NIH_ZGC_7 Danio rerio cDNA clone IMAGE:7052472
 DEFINITION 5', mRNA sequence.
 ACCESSION CK028187
 VERSION CK028187.1 GI:38554111
 KEYWORDS EST:
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 1117)
 NIH-MGC <http://mgs.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Len Zon, Harvard
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LAM14826 row: j column: 22
 High quality sequence stop: 835.
 Location/Qualifiers
 1. 1117
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:7052472"
 /tissue_type="whole body"
 /lab_host="DH10B"

FEATURES
 source

```

/clone_lib=NIH_ZGC-7"
/notes=Vector: pExpress1; Site.1: NotI; Site.2: EcoRV;
Bulk tissue was collected from a whole adult individual
from the Tuebingen strain. 1st strand cDNA was primed with
a Not I - oligo(dT) primer, double-stranded cDNA was
cloned into the Not I and EcoRV sites of pExpress-1.
Library was size-selected for >1 kb fragments and
normalized. A non-normalized version of this library is
also available (NIH_ZGC-0). Library was constructed by
Open Biosystems (Huntsville, AL)."

```


Alignment Scores:

Pred. No.: 2.15e-20 Length: 946
 Score: 682.00 Matches: 165
 Percent Similarity: 52.59% Conservatives: 28
 Best Local Similarity: 44.96% Mismatches: 75
 Query Match: 32.12% Indels: 99
 DB: 14 Gaps: 11

US-10-087-080-32 (1-402) x CD754754 (1-946)

```

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db 76 ATGAACCTGGAGGTTTCTGCG-----GGGGTCACCTACCGCTCCAGCTCGGAGCTG 129

Qy 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
Db 130 TGCAGCGATGCTGAGGCGAGCATCCCATCGCGGTGTCGCA-----GAGGAGGAGCTG 193

Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60
Db 184 GGCTCGGATGGAGACTGGGTGGCGCACAGTCCGCA-----CCTGTCCGCGACACCAA 237

Qy 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAla 80
Db 238 GGC-----

Qy 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
Db 240 -----

Qy 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
Db 241 -----AAACCCCTACACTCGGAGACCCAAACCTCCA 270

Qy 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
Db 271 TACTCTTACATCGGCACCTATCGCATGGCCATCGAGACTCCAACTCCGCGCGACTCACT 330

Qy 141 LeuAlaGluIleAsnGlyTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
Db 331 CTAGCCGAATCAACGACTACCTCATGAGAGTTCCTCGTTTATAGAGGCAGCTACACC 390

Qy 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
Db 391 GGCTGGAGGAACTCAGTGGCCCAATACTCTCTAAACGACTGCTTTCTCAAGGCTTA 450

Qy 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
Db 451 AGGATCTCTCGACCGTGGGAAAGACAAATTACTGGATGCTGAACCCGACAGCGAG 510

Qy 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSerHisArgAlaPro 220
Db 511 TACACCTTCGGGACGGAGTGTTCGAGAGGAGAAAGCGCATTAGTAAATAAACCC--- 567

Qy 221 ValProAlaProGlyLeuArgProGluAlaProGlyLeuProAlaAlaProProPro 240
Db 568 -----GGCCCGGTGCCAA 594

Qy 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArgAla 260
Db 595 ACCCAGCATGGACAGCAATGACTCCATCGCTACGCT----- 633

Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
Db 634 CCTTCGAGCGGANNAGTTCACAGTTCCTTCGCATTGAGAGCATCTTCAGTCAGCCCTTC 693

Qy 281 ArgSerArgArgLeuArgAsp-ThrAlaProGlyThrThrLeuGlnThrProGlyAlaAlaPr 300
Db 694 AGA-----AGGAGGACCGCCCGTGTCTGACCCCTGACACCTGGCGGGGGA 741

Qy 300 o-----CysProProLeuProAlaPheProAlaLeuLeuProAlaAla----- 314
Db 742 GTGGACACTGTGCGCGCGCTATGTCATGCTCGATCTTGTGTCCTTGTGTCCTTGNAGACCCNNNC 801

```

```

Qy 315 -----ProCysArgAlaLeuLeuProLe 322
Db 802 NNGAGNCGCGANGCGCGACTTTTTCGCATTTCAGTCGCGCTGGGATGCTCTCATCNG 861
Qy 322 uCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluValProPr 342
Db 862 NTACNAGCGCGCTGCGCGCACCTGCACGCGGGATTCCACGTCATNTGATATCTGCC 921

Qy 342 oThrAlaProProLeuLeu 348
Db 922 NACN-----CGGTTACTC 934

```

RESULT 4

CF997203

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF997203 799 bp mRNA linear EST 25-NOV-2003
 AGENCOURT_16391898 NIH_ZGC_7 Danio rerio cDNA IMAGE:7036173
 5', mRNA sequence.

CF997203

CF997203.1 GI:38518054

EST

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-remail.nih.gov

Tissue Procurement: Len Zon, Harvard

cDNA Library Preparation: Open Biosystems

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Place: LLNL14784 row: c column: 19

High quality sequence start: 10

High quality sequence stop: 715

Location/Qualifiers

1..799

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:7036173"

/tissue_type="whole body"

/lab_host="DH10B"

/clone_lib="NIH_ZGC_7"

/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;

Bulk tissue was collected from a whole adult individual

from the Tuebingen strain. 1st strand cDNA was primed with

a Not I - oligo(dT) primer, double-stranded cDNA was

cloned into the Not I and EcoRV sites of pExpress-1.

Library was size-selected for >1 kb fragments and

normalized. A non-normalized version of this library is

also available (NIH_ZGC_10). Library was constructed by

Open Biosystems (Huntsville, AL)"

ORIGIN

Alignment Scores:

Pred. No.: 3.24e-20 Length: 799

Score: 676.00 Matches: 157

Percent Similarity: 54.88% Conservatives: 23

Best Local Similarity: 47.87% Mismatches: 52

Query Match: 31.84% Indels: 96

DB: 14 Gaps: 11

US-10-087-080-32 (1-402) x CF997203 (1-799)

QY 1 MetLeuLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 DB 85 ATGAACCTGGAGTTTCTTGC-----GGGAGTCACTACGACTCCAGCGCTGGGAGTTG 138
 QY 21 GluGlyAlaGlyClySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
 DB 139 TGCAGCGATGCTCAGGGGAGCATCCCATCGCGGTGTCGCA-----GAGGAGAGCTG 192
 QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60
 DB 193 GACTCGATGGAGACTGCTGGCGCACAGTCCGCA-----CCTGTGGCGACACCAAA 246
 QY 61 GlyAspGlyGluClnSerAlaGlyGlyProGlyAlaGluGluAlaLeuProAlaAla 80
 DB 247 GGC----- 249
 QY 81 AlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
 DB 249 ----- 249
 QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
 DB 250 -----AAACCCTACACTCGGAGACCCCAAACTCCA 279
 QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleAlaArgAspSerAlaGlyGlyArgLeuThr 140
 DB 280 TACTCTTACATCGCACTTATCGCCATGCGCATCCGAGACTCCCACTCGCGCGACTCACT 339
 QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
 DB 340 CTAGCCGAATCAACGACTACCTCATGAAGAAGTCCCGTTTTTTAGAGCGCAGCTACACC 399
 QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
 DB 400 GGCTGGAGAACTCAGTGGCGCAATATCTGTCTCTAAACGACTGCTTCTTAAGTCTTA 459
 QY 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
 DB 460 AGGATCTCTCGAGACCGTGGGAAGGACATTAATCTGATCTGAACCGCACAGCGAG 519
 QY 201 TyrThrPheAlaAspGlyValPheArgArgArgArgLysArgLysArgLysHisArgAlaPro 220
 DB 520 TACACCTTCGCGAGCGAGTGTTCAGAGAGAGAGAGAGCGCATTAAGTAAATAAACCC--- 576
 QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240
 DB 577 -----GGCAGGGAGCCAGAG-----GGGCGGTGCNA 603
 QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaAla 260
 DB 604 ACCACGCAATGGACAGCAATGACTCCTCATCGCTACGCT----- 642
 QY 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
 DB 643 CCTTCAGCGGANAGATCCACAGTCTTTCGCAATGAGAGATCCTCAGTCAACCCCTTC 702
 QY 281 ArgSerArgArgLeuArg-----AspThrAlaProGlyThrThrLeuGln 295
 DB 703 AGA---AGGAGGACCGCCCGTGTCTACGCCCTGACACCTGGCGGGG----- 747
 QY 296 TrpGlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaPro 315
 DB 748 ---GGGAGTGGACACTGT-----GCTGCCGTA 771
 QY 316 CysArgAlaLeuLeuProLeuCys 323
 DB 772 TGTCATGCTTCGATCCTATGCTGC 795

RESULT 5
 A1169632/c
 LOCUS

559 bp mRNA linear EST 20-JAN-1999

DEFINITION EST215519 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
 RKIBV46 3' end, mRNA sequence.

ACCESSION A1169632
 VERSION A1169632.1 GI:3709672
 KEYWORDS EST.
 SOURCE Rattus sp.

ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 559)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kervilave,A.R. and Adams,M.D.

AUTHORS Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index

JOURNAL Unpublished (1998)
 COMMENT Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org

Seq primer: M13-21.
 Location/Qualifiers
 1..559

source
 /organism="Rattus sp."
 /mol_type="mRNA"
 /db_xref="ATCC (inhost):2027517"
 /db_xref="taxon:10118"
 /clone="RKIBV46"
 /clone_lib="Normalized rat kidney, Bento Soares"
 /note="Organ: kidney; Vector: pT7T3pac; Site_1: EcoRI;
 Site_2: NotI"

ORIGIN
 Alignment Scores:
 Pred. No.: 1..53e-19 Length: 559
 Score: 656.50 Matches: 138
 Percent Similarity: 76.88% Conservative: 5
 Best Local Similarity: 74.19% Mismatches: 40
 Query Match: 30.92% Indels: 3
 DB: 9 Gaps: 3

US-10-087-080-32 (1-402) x A1169632 (1-559)
 QY 198 AsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLysHis 217
 DB 557 AACAGCGAATACACTTCGCGCGGCTCTTCGCGCGCGCGCAAGCGCTTCAGCCAC 498
 QY 218 ArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
 DB 497 CGGACCACAGTCTCCGATCGGGCTACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 438
 QY 237 AlaProProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGln 256
 DB 437 ACCCGCAGCGCGCGCCACCGCGGCTCTCCCATCGCGGCTCGCGGCTCGCGG 378
 QY 257 GluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeu 276
 DB 377 GAGGAGGGCTCCAGCGCGGAGCAAGTCTCCAGCTCTTCGCGCATCGCAGCATCCTC 318
 QY 277 ArgLysProPheArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrp 296
 DB 317 AGCAAGCGGTTTCGACCGCGCGCGGAGCGCGGCTCTGGGGGTGCGAGTACCTCG 258
 QY 297 GlyAlaAlaProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCys 316
 DB 257 AGCGTGTCTCCTCGCGCGGCTCGCGGCTATCCCGGCTCTTCCCGGCTCGTCCGGC 198
 QY 317 ArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArg 336
 DB 197 GGTGCCCTGTCTGCTGCTTACGGCGCGGCGGAGCCCGCGCTGTGTGGCTCGGCG 138

337 GluAlaGluValProThrAlaProProLeuLeuAlaProLeuProAlaAlaAla 356
 137 GGGGCCGAGGTGAACCCGGGGG---CCCTGTGCTGGCGCCCTCTCCACCGGGCC 81
 357 ProAlaLysProLeuArgGlyPro---AlaAlaGlyGlyAlaHisLeuTyrCysProLeu 375
 80 CCAGCAAGCCATTTCGAGTCCGGAGACCGCGCGCGCGGCGCACCTGTACTGCCCCCTA 21
 376 ArgLeuProAlaAlaLeu 381
 20 CGGCTGCCACCGGCCCTG 3

RESULT 6
 CK030218 804 bp mRNA linear EST 26-NOV-2003
 LOCUS AGENCOURT_16624545 NIH_ZGC_7 Danio rerio cDNA clone IMAGE:7051925
 DEFINITION 5', mRNA sequence.
 ACCESSION CK030218
 VERSION CK030218.1 GI:38556142
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 804)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Len Zon, Harvard
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM14825 row: d column: 03
 High quality sequence stop: 716.

FEATURES
 source
 1..804
 Location/Qualifiers
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7965"
 /clone="IMAGE:7051925"
 /tissue_type="whole body"
 /lab_hosts="DH10B"
 /clone_lib="NIH_ZGC_7"
 /notes="vector: pExpress1; Site 1: NotI; Site 2: EcoRV;
 Bulk tissue was collected from a whole adult individual
 from the Tuebingen strain. 1st strand cDNA was primed with
 a Not I - oligo(dT) primer, double-stranded cDNA was
 cloned into the Not I and EcoRV sites of pExpress-1.
 Library was size-selected for >1 kb fragments and
 normalized. A non-normalized version of this library is
 also available (NIH_ZGC_10). Library was constructed by
 Open Biosystems (Huntsville, AL)"

ORIGIN
 Alignment Scores:
 Pred. No.: 1..41e-17 Length: 804
 Score: 615.00 Matches: 141
 Percent Similarity: 58.36% Conservative: 16
 Best Local Similarity: 52.42% Mismatches: 40
 Query Match: 28.97% Indels: 72
 DB: 14 Gaps: 8

US-10-087-080-32 (1-402) x CK030218 (1-804)

QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 DB 95 ATGAAGTTGGAGGTTTC-----TCTCGGAGTCCGCTTGGGACAGCCCTCGACCTG 148
 QY 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
 DB 149 TGCAGTGACATGGATGTGAACGCTTCCCGCTGTCCACG-----GAGGAGGAATTG 202
 QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaArgAspThrGln 60
 DB 203 GGCTCGGATGGAGACTGCTCAGCCAAC----- 229
 QY 61 GlyAspGlyGluGlnSerAlaAlaGlyGlyProGlyAlaGluGluAlaAlaProAlaAla 80
 DB 230 -----AGTCCAGGACCGGTGCCCT----- 250
 QY 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
 DB 251 -----GTTCCGGACGGG----- 262
 QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
 DB 263 -----AGGCGAAACCATATACGCGTAGACCCCAACACCT 298
 QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
 DB 299 TACTCTTACATTGCACCTATTTCGAATGGCCATACGCGATTCCAAACACCGGTGCTTTACA 358
 QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArgGlySerTyrThr 160
 DB 359 CTTCCTGGAATAAAGCAATACTTAATGAAAAAATCCCGTTTTTCGGGCGAGCTACACG 418
 QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
 DB 419 GGGTGGCGAAACTCTGTACGCATAAATCTGTCTATTAACGACTGCTTTTAAAAAGTTTG 478
 QY 181 ArgAspProSerArgProTropGlyLysAspAsnTyrTrpMetLeuAsnProAsnSerGlu 200
 DB 479 CGAGATCCCTCGCTCGCTGGGCGAAGACATATTATTGGATGTTTGAACCCACACAGCGAG 538
 QY 201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysArgLysArgAlaPro 220
 DB 539 TACACTTTTCAGACGGCGGTTTTCCCGCGGAGGAGGAAGCGCATCAGT----- 586
 QY 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240
 DB 587 -----AGAGAGATCTCGGCGAGCGCGGGAATCCCCAGAGCGC 622
 QY 241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluAlaAla 260
 DB 623 GTGCCCGCGGATGACAGC-----AGGCTTCCAGCGCGGCGAGGAAGC----- 664
 QY 261 SerProAlaGlyLysPheSerSerSer 269
 DB 665 -----GTCTCTAAAGTTTTCAGAGCTCT 685

RESULT 7
 BQ285521 581 bp mRNA linear EST 14-MAY-2002
 LOCUS faa51g09.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone
 DEFINITION IMAGE:5912369 5', similar to SW:HPH1_RAT Q83244 HEPATOCYTE NUCLEAR
 FACTOR 3 FORKHEAD HOMOLOG 1 ; mRNA sequence.

ACCESSION BQ285521
 VERSION BQ285521.1 GI:20654021
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 581)
 Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M.,
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R.,
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R. and Wilson, R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@watson.wustl.edu
 CDNA Library Preparation: Raymond Lee. CDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 Resourcenet/Primarbatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 487.

FEATURES

Location/Qualifiers

1..581

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone_image="5912369"

/sex="mixed male and female"

/issue_type="3 day fin regenerates"

/lab_host="E. coli XL0LR"

/clone_lib="zebrafish fin day3 regeneration"

/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st

strand cDNA primed with (GA)10ACTAGTCGAG(T)18, followed

by second strand synthesis, and ligated to 5' adapter

(5')-aattgcgcagag-3', 3'-gccgtgtc-5'. cDNA was cloned

directionally (EcoRI/XhoI) into Stratagene Zap express

lambda phage arms. Mass invivo excision done to obtain

inserts in pBK-CMV phagemid."

ORIGIN

Alignment Scores:

Pred. No.: 1..09e-17 Length: 581
 Score: 614.00 Matches: 129
 Percent Similarity: 64.41% Conservative: 14
 Best Local Similarity: 58.11% Mismatches: 26
 Query Match: 28.92% Indels: 53
 DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ285521 (1-581)

QY 109 ArgSerLysProTyrThrArgArgProLysProTyrSerTyrLeuAlaLeuLeuAla 128
 DB 35 AAAGGCAACCCCTACACTCGGAGACCAACCTCCATCTTATCATCGCATTCGCC 94
 QY 129 MetAlaIleArgAspSerAlaGlyArgLeuThrLeuAlaGluLeuAsnGluTyrLeu 148
 DB 95 ATGGCCATCCGAGACTCCAACTCCGCGCGACTCACTAGCGGAAATCAAGACTACCTC 154
 QY 149 MetGlyLysPheProPheArgGlySerTyrThrGlyTyrArgAsnSerValArgHis 168
 DB 155 ATGAAGAAGTCCCGTTTTTATAGACGACGTACACCGCTGGAGGAATCTAGTGGCCAT 214
 QY 169 AsnLeuSerLeuAsnAspCysPheValLysValLeuAtgAspProSerArgProTyrGly 188
 DB 215 AATCTGCTCTAAACGACTGCTTCTCAGGCTTTAAGGGATCCTTCGAGACCGGGGGA 274
 QY 189 LysAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPhe 209
 DB 275 AAGGCAAAATCTAGTGTGTAACCGCGACGAGTACACCTTCGCGACGGAGTGT 334
 QY 209 ArgGArgArgGlyArgLeuSerHisArgAlaProValProAlaProGlyLeuArgPro 228

DB 335 CGCAGAGGAGAAAGCGCATTAGTAAAAAAC-----GGCAGGAGGCCA 379
 QY 229 GCGluAlaProGlyLeuProAlaAlaProProAlaProAlaAlaProAlaAspPro 248
 DB 380 GAG-----GGCCCGTGCACCAACCCAGCATTTGGACAGCAATGAC 418
 QY 249 ArgMetArgSerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSer 268
 DB 419 TCCATCGTACGCT-----CCTTCGAGCGGAAAGTTCCACCACT 457
 QY 269 SerPheAlaLeuAspSerIleLeuArgLysProPheArg-SerArgArgLeuArgAspTh 288
 DB 458 TCCTTCGCCATTGAGAGCATCCTCAGTCGACCTCTCAGAGGGAGG----- 503
 QY 288 rAlaProGlyThrThrLeuGlnTrpGlyAlaAlaProCysProProLeuProAlaPhePr 308
 DB 503 ----- 503
 QY 308 oAlaLeuProAlaAlaProCysArgAlaLeuLeuProLeuCysAlaTyrGlyAlaGl 328
 DB 504 -----ACCGCCCGTGTCTCAGCCCTGACACCT-----GCCCGGGG 538
 QY 328 YGlu 329
 DB 539 GGAG 542

RESULT 8

EX084216

LOCUS

DEFINITION

EX084216 AGENAE Rainbow trout normalized multi-tissues library

(tcac) Oncorhynchus mykiss cDNA clone tcac0005c.1.17 5prim, mRNA

sequence.

EX084216

VERSION

EX084216.1 GI:27747633

KEYWORDS

EST.

SOURCE

Oncorhynchus mykiss (rainbow trout)

ORGANISM

Oncorhynchus mykiss

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

1 (bases 1 to 805)

Govoroun, M., Guiguen, Y. and Le Gac, F.

Construction and primary characterization of normalized cDNA

libraries in rainbow trout, Oncorhynchus mykiss

Unpublished (2003)

Contact: Guiguen Y

INRA - SCRIBE

Campus de Beaulieu, RENNES cedex, 35042, France

Tel: 02.23.48.50.09

Fax: 02.23.48.50.20

Email: Yann.Guiguen@beaulieu.rennes.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us

at sigenasupport@jouy.inra.fr to obtain the chromatogram of this

sequence.

Plate: 0005 row: 1 column: 17

Seq primer: T7.

Location/Qualifiers

1..805

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:8022"

/clone="tcac0005c.1.17"

/tissue_type="adipose tissue, blood, brain,

differentiating gonads, gills, interrenal, intestine,

kidney, liver, muscle, ovary, pituitary, testis"

/dev_stage="from embryos to adults"

/lab_host="DH10B"

/clone_lib="AGENAE Rainbow trout normalized multi-tissues

library (tcac)"

/notes="vector: pT73D-pac; Clone distribution: AGENAE

Resource Centre, Francois PIUMI,

Francois.PIUMI@jouy.inra.fr, INRA, CEA Radiobiologie et

Etude du genome (LREG), Domaine de Vilvert, 78352,

ORIGIN
Jouy-en-Josas cedex, France"

Alignment Scores:
Pred. No.: 1.64e-17 Length: 805
Score: 613.50 Matches: 143
Percent Similarity: 56.43% Conservative: 15
Best Local Similarity: 51.07% Mismatches: 43
Query Match: 28.90% Indels: 80
DB: 13 Gaps: 9

US-10-087-080-32 (1-402) x BX084216 (1-805)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGly----- 17
Db 174 ATGAACCTTGAAGTATCTGTGGAAC---CACCAACAGGAGTGAATCTTTCAGGAG 230
Qy 18 -----SerAspLeuGluGlyAlaGlySerAspAlaProSerProLeuSerAlaAla 35
Db 231 CTGTGCAGCGATCCGAGGGGAGC-----GTCCGCTCGCTCTCTCC----- 272
Qy 36 GlyAspSerLeuGlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGly 55
Db 273 GGTGAGGAGAGTGGGTTCGATGGGACTGCTGGG----- 311
Qy 56 AlaArgAspThrGlnGlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGlu 75
Db 312 -----CAC 314
Qy 76 AlaIleProAlaAlaAlaAlaValAlaGluGlyAlaGluAlaGlyAlaAla 95
Db 315 AGCCGCCACCTGTGCAC----- 332
Qy 96 GlyProGlyAlaGlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArg 115
Db 333 ---CCCGCTAGACGGCATAGGATTGGC-----AAACCGTATATACGG 374
Qy 116 ArgProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAla 135
Db 375 AGACAAACCTCCGTACTCTTACATCGACTGATGCCATGATCCGATCGGTACTCAACA 434
Qy 136 GlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePhe 155
Db 435 TCTGCTGCTGACTCTGGCAGAGATCAATGACTATCTGATGAAAGAGTTTCCATTTTC 494
Qy 156 ArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCys 175
Db 495 CGGGCAGCTACACTGGTTGGAGAAATTCGGTGGCCGACCAACTTGTCACTTAACGACTGT 554
Qy 176 PheValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTrpMetLeu 195
Db 555 TTCTTAAGGTCCTCCGGGACCCGTCAGACCCCTGGGMAAAGACAACTACTGATGCTG 614
Qy 196 AsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeu 215
Db 615 AATCCTCAGCAGGTACACCTTCGTCAGCGAGTGTTCGACGAGAGAAACGCGATC 674
Qy 216 SerHisArgAlaProValProAlaProGlyLeuArgPro-----GluGluAla 231
Db 675 AATAAAGACAGG---CAAGAGCAAAACGTCCTCCGACCATCTCCGACGAGACGCGCA 733
Qy 232 Pro-GlyLeuProAlaAlaProProAlaProAlaProAlaProAlaProAlaSerProArgMetAr 251
Db 734 AGTCACCATCATCATCCACCCCTGTCACATAAAGCGNACCGACGAT----- 782
Qy 251 gSerProAlaArgGlnGluGluArgAlaSerProAlaGlyLysPheSerSerSerPhe 270
Db 783 -----GCCAAGTCTCCAGTTCCTTC 803

RESULT 9
BQ449254
LOCUS faa45a03.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone
DEFINITION

IMAGE:5911949 5', similar to TR:070220 070220 FORK HEAD
TRANSCRIPTION FACTOR. ; mRNA sequence.

ACCESSION BQ449254
VERSION BQ449254.1 GI:21252366
EST:
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 564)
Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M.,
Edg, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 450.
Location/Qualifiers
1..564
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5911949"
/sex="mixed male and female"
/tissue_type="3 day fin regenerates"
/lab_host="E. coli XL0LR"
/clone_lib="zebrafish fin day3 regeneration"
/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st
strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed
by second strand synthesis, and ligated to 5' adapter
(5'-aactggcagag-3', 3'-gccgtgcc-5'. cDNA was cloned
directionally (EcoRI/XhoI) into Stratagene Zap express
lambda phage arms. Mass in vivo excision done to obtain
inserts in pBK-CMV phagemid."

ORIGIN
Alignment Scores:
Pred. No.: 1.42e-17 Length: 564
Score: 611.00 Matches: 128
Percent Similarity: 61.14% Conservative: 12
Best Local Similarity: 55.90% Mismatches: 29
Query Match: 28.78% Indels: 60
DB: 13 Gaps: 5

US-10-087-080-32 (1-402) x BQ449254 (1-564)

Qy 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db 9 ATGAACCTTGAAGTATCTGTGGAAC-----GGGGTCTACTAGCACTCCAGGCTCGGAGCTG 62
Qy 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
Db 63 TGACGCGATGCTGAGGGGAGCATCCCATCCCGGTGTCGCA-----GAGGAGGAGCTG 116
Qy 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaArgAspThrGln 60

```

Db      117 GCCTGGATGAGACTGGTGGCGCACAGTCCGGA-----CCTGTGCGGCACACCAA 170
Qy      61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaProAlaAla 80
Db      171 GGC-----173
Qy      81 AlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
Db      173 -----173
Qy      101 GlyAlaGlySerGlyGluGlyAlaArgSerIysProTyrThrArgArgProIysProPro 120
Db      174 -----AAACCTACACTCGAGACCCAAACCTCCA 203
Qy      121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyValArgLeuThr 140
Db      204 TACTTTTATGTCGCACTTATGCGCATGCGCATCCGAGACTCAACTCCGCGGCACTACT 263
Qy      141 LeuAlaGluIleAsnGluTyrIleuMetGlyLysPheProPheArgGlySerTyrThr 160
Db      264 CTAGCGCAAAATCAACGCACTACCTCATGAAGAAGTTCCTGTTTATAGAGCGAGCTACACC 323
Qy      161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeu 180
Db      324 GCGTGGAGAACTCAGTGGCGCATTAATCTGTCTCTAAACGACTGCTTTCTCAAGGTCTTA 383
Qy      181 ArgAspProSerArgProTyrIleuMetGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
Db      384 AGGGATCTCTTCGAGACCGTGGGAAAGGACAACTACTGATGCTGAACCCGACAGCGAG 443
Qy      201 TyrThrPheAlaAspGlyValPheArgArgArgLysArgLysArgLysArgLysArgLysArgLys 220
Db      444 TACACCTTCGCGAGCGGAGTGTTCGCAAGAGGAGAAAGCGCATTTAGTAAAGAAAC--- 500
Qy      221 ValProAlaProGlyLeuArgProGlu 229
Db      501 -----GGCAGGAGGCCAGAG 515

```

RESULT 10

BF282916/c

LOCUS

DEFINITION EST447507 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

↑ACCESSION BF282916 GI:11213986

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 514)

Malek, R.L., Cho, J., Lee, Y., Karamycheva, S., Parvizi, B., Pertea, G.,

Sultana, R., Teal, J., White, J., Quackenbush, J. and Lee, N.H.

Generation of ESTs from Normalized Rat Embryo, Bento Soares

Unpublished (2000)

Other ESTs: EST348098

Contact: Lee, NH

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC

tel#703-385-2700 for further information.

Location/Qualifiers

1..514

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="RGIDU92"

/tissue_type="mixed tissue"

FEATURES

source

/lab_host="DH5-alpha"

/clone_lib="Rat Gene Index, normalized rat, Rattus

norvegicus cDNA"

/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;

Combination of ROV, RBR, RLI, RPL, RLU, REM, RMO,

RSP, RHE, RPC, RPN"

ORIGIN

Alignment Scores:

Pred. No.: 1..47e-16

Score: 586.50

Percent Similarity: 75.58%

Best Local Similarity: 72.67%

Query Match: 27..63%

DB: 10

Gaps: 3

US-10-087-080-32 (1-402) x BF282916 (1-514)

Qy 213 LysArgLeuSerHisArgAlaProValProAlaProGlyLeuArgProGlyLeuAlaPro 232

Db 513 AAGCGGCTCAGCCACCGGACACAGTCTCCGATCGGGCTACGGCGGAGGAGCCCA 454

Qy 233 GlyLeuProAla--AlaProProAlaProAlaProAlaProAlaSerProArgMetArg 251

Db 453 CCGGACCTCGGGGACCCCGAGCCGCCACCCCGGGTCTTCCCAATCGCGCGC 394

Qy 252 SerProAlaArgGlnGluArgAlaSerProAlaGlyLysPheSerSerPheAla 271

Db 393 TCGCCGCTCGCCAGGAGGAGGCTCCAGCCGCGGAGCAAGTCTCCAGTCTCTCGCC 334

Qy 272 IleAspSerIleLeuArgLysProPheArgSerArgArgLeuArgAspThrAlaProGly 291

Db 333 ATCAGACAGATCTCTCAGCAAGCGGTTTCGAGCCGCGCGGAGCGGCGCGGCTTGGGG 274

Qy 292 ThrThrLeuGlnTyrGlyAlaAlaProCysProProLeuProAlaPheProAlaLeu 311

Db 273 GTCAGCTACCTGGAGCGCTGCTCTCCGCGCGCTGCGCGCTATCCCGGCTCTT 214

Qy 312 ProAlaProCysArgAlaLeuLeuProLeuProLeuProLeuProLeuProAla 331

Db 213 CCGCGCTCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154

Qy 332 ArgLeuGlyAlaArgGluAlaGluValProProThrAlaProProLeuLeuAlaPro 351

Db 153 CTCTGCGCTCGCGCGGCGCGGAGGTGCAACCGCGCGG---CCCCTGTTGCTGCGCGCC 97

Qy 352 LeuProAlaAlaAlaProAlaLysProLeuArgGlyPro---AlaAlaGlyAlaHis 370

Db 96 CTCTCCACGCGCGCGCGCGAGCCATTTCGAGGTTCGAGAGTCCGAGAGCCCGCGCGCGCAC 37

Qy 371 LeuTyrCysProLeuArgLeuProAlaAlaLeuGln 382

Db 36 CTGTACTGCCCCCTACGGCTGCCACGCGCGCTGCAG 1

RESULT 11

AL636071

LOCUS

DEFINITION

AL636071 XGC-neurula Silurana tropicalis cDNA clone TNeu015112 5',

mRNA sequence.

AL636071

VERSION

KEYWORDS

SOURCE

ORGANISM

Silurana tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Silurana.

1 (bases 1 to 682)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11_2003)

Unpublished (2003)

On Nov 7, 2001 this sequence version replaced gi:16788050.

Contact: Huckle E

Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: tropesanger.ac.uk
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
 cDNA was oligo dt primed from Sug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
 Vector: pCS107, Site1: EcoRI; Site2: NotI
 Host: Escherichia coli DH10B
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS SEQUENCE ID: TNeu015112.plkSP6
 Sequencing primer: SP6

FEATURES
 source
 1..682
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TNeu015112"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /clone_lib="XGC-neurula"
 /notes="Vector: pCS107; Site1: EcoRI; Site2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,11e-16 Length: 682
 Score: 586.00 Matches: 129
 Percent Similarity: 61.51% Conservativeness: 18
 Best Local Similarity: 53.97% Mismatches: 50
 Query Match: 27.60% Indels: 42
 DB: 9 Gaps: 6

US-10-087-080-32 (1-402) x AL636071 (1-682)

Qy 62 AspGlyGluInSerAlaGlyGlyProGlyAlaGluGluAlaLeProLaAlaLa 81
 Db 7 GACACTGACACCGGTAGCTTGGGGGGATGAAGAGTAA----- 48
 Qy 82 AlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGlyGly 101
 Db 49 -----GTGGAAGAAGAGGAGGAGGATTAACCCAGAGAGAAATGGGTCTCGCA 96
 Qy 102 AlaGlySerGly-----GluGlyAlaArgSerLysProTyrThr 114
 Db 97 GATGTTCTACGCAATGCCGAGCACAAATTTGTTAGGGGGCAACAAAGACTTACACC 156
 Qy 115 ArgArgProLysProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSer 134
 Db 157 CGCCTTTGGAACCTCTTACTCTCATCGCACTCATGCCATGCTATTAAAGACTCA 216
 Qy 135 AlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhe 154
 Db 217 GCAGTGGGGCGCTGACCCCTTGAGAGATCAATGACTATCTGATGAAGATTCCCATTC 276
 Qy 155 PheArgGlySerTyrThrGlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsp 174
 Db 277 TTCAGGGGTAGTACACGGCTGGAGAACTCGGTGCGACACAATCTTTCCCTCAACGAC 336
 Qy 175 CysPheValIysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTpMet 194
 Db 337 TGTTTCGTCAAGTTCTGGCGGACCCCATCAGCCCTTGGGGGAAAGACAACTACTGGATG 396
 Qy 195 LeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArg 214
 Db 397 CTGAACCCGAATAGTAGTACCTTCGACAGCGGCTTTCAGACGCAAGAGGAGCGCA 456
 Qy 215 LeuSerHisArgAlaProValProAlaProGlyLeuArgProGluGluAlaProGlyLeu 234
 Db 457 CTTAACAGGGTCAACCAAGTGC-----CTCAAGAGCAGGATTTGCAGAGGCTT 504

Qy 235 Pro-----AlaAlaProPro 239
 Db 505 GCAGAACAGCAGCATCAATAATGATGAACCCCAAAAGCCAGCCAGGATCTTCTCCCTCA 564
 Qy 240 Pro-----AlaProAlaAlaProAlaSerProArgMetArgSerProAla 254
 Db 565 TCGAGCGGCTGATCATGGCACCCTCAGCACCATCATCTCTCTACAAACTCTCTCTCC 624
 Qy 255 ---ArgGlnGluGluArgAlaSerProAlaGlyLysPheSerSerPheAlaIle 272
 Db 625 AACAGGTCCTGCCAAAGAAACCACTCANGGACCAAGTTTCCAGCTCTCTTGCATT 681

RESULT 12
 CNS044M2
 LOCUS
 DEFINITION
 Tetraodon nigroviridis genome survey sequence T7 end of clone
 081F22 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION
 AL274259.1 GI:7996530
 VERSION
 GSS: genome survey sequence.
 KEYWORDS
 Tetraodon nigroviridis
 SOURCE
 Tetraodon nigroviridis
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
 1 Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence
 Nat. Genet. 25 (2), 235-238 (2000)
 2 20296633
 PUBMED
 10835645

TITLE
 Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence

JOURNAL
 MEDLINE
 PUBMED
 10899143

REFERENCE
 3 (bases 1 to 878)
 Direct Submission
 Genoscope.
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source
 1..878
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="081F22"
 /clone_lib="G"
 /note="Genoscope sequence ID : COB081DC11LP1-end : T7"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,03e-15 Length: 878
 Score: 566.00 Matches: 144
 Percent Similarity: 52.48% Conservativeness: 15
 Best Local Similarity: 47.52% Mismatches: 77
 Query Match: 26.66% Indels: 68
 DB: 29 Gaps: 9


```

US-10-087-080-32 (1-402) x CNS044M2 (1-878)

QY 30 SerProLeuSerAlaAlaGlyAspAspSerLeuGlySerAspGly----- 44
Db 95 TCGCCAAACAGCGCG-----TCGTGGGTTTCATGTGTCCTGACGCGCAGAGA 142
QY 45 -----AspCysAlaAlaAlaGlyPro 50
Db 143 CTTTITAGATGCCATATTGTGGCGCTGCTGCCGTAGAGGAGGACATCGTATTAAACAT 202
QY 51 SerAlaGlyGlyAlaAlaG-----AspThrGlnGlyAspGlyGluGln 65
Db 203 GAAGCTGAGGTGTGTGTCGGCGGCGCACTATGATAAAAAACACGAGGTGACCGGAGMG 262
QY 66 SerAlaGlyGlyGlyPro---GlyAlaGluGluAlaIleProAlaAlaAlaAlaAla 84
Db 263 GGAGGAGGAGCVCT-CCCTGTCTGTTGAAGAGCGCTGGCGTCASTAGGAGCTRCGTG 321
QY 85 ValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGlyGlyAlaGlySer 104
Db 322 GTGGCGCASCAGCGAGCTGTGACTCCATCGCGC----- 354
QY 105 GlyGluGlyAlaAlaSerIleProThrArgProGlyProGlyProGlyProGlyProGly 124
Db 355 -----GTGGAGCGCCAGAGTAAACCGTACATACGAGACCTAACCTCTCTCTACATC 411
QY 125 -AlaLeuIleAlaMetAlaIleAlaArgAspSerAlaGlyArg-LeuThrLeuAlaGluI 144
Db 412 AGCCCTCATCGCATGTCGCAATWCGGAGCTCCGGTACTGGCAGCGCTGATTCGGCGGAGA 471
QY 144 leAsnGluTyrLeuMetGlyIlePheProPheArgGlySerTyrThrGlyTrpAlaG 164
Db 472 TCAACAACATPACTGATGACAGAAATTTCCATTTTGTAGAGGAGCGACACCGGCTGGAGGA 531
QY 164 snSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArgAspProS 184
Db 532 ACTCTGTTGACACAACTGCTCTTAACGAGCTGCTTCTCAAGTGCTCCGCGACCGGT 591
QY 184 eArgProTrpGlyIleAspAsnTyrTrpMetLeuAsnProAsnSerGluTyrThrPheA 204
Db 592 CCAGACCTTTGGGGAAGAGCAATTAAGTGTCTGAATCCCAACAGTACAGTATACCTTTK 651
QY 204 laAspGlyValPheArgArgArgGlyAlaGlyLeuSerHisArgAlaProValProAlaP 224
Db 652 CGATGGAGTGTTCCGACCGAGAGAGCGGATTCACAAAAA----- 695
QY 224 roGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProAlaProAlaA 244
Db 696 -----CTAAACACGGAG-----CCAACGGCGCGAG 720
QY 244 laProAlaSerProArgMetArgSerProAlaArgGlnGluGluAlaSerProAlaG 264
Db 721 AGCAGCARGAGCGACGGCATTTTCCTGTCACGAGACCGAT-----TCGTGCG 771
QY 264 lYlPheSerSerPheAlaIleAspSerIleLeuArgIlyProPheArgSerArgA 284
Db 772 GSAAGTWTAGAGTACATTACAGWAGACAGCATCTCAGCAAGCCCTTCAAG----- 824
QY 284 rgLeuArgAspThrAlaProGlyThrLeuGluTyrTrpGlyAlaAlaProCysProProL 304
Db 825 -----AGGAGCCCGCAGACACCATTTGTGAGCCCGCTCCTGCTTCACTGCGCA 873
QY 304 euPro 305
Db 874 GCGCC 878

RESULT 13
AI586081
LOCUS
DEFINITION
v92c05.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1209608 3, similar to gb.U36760 Mus musculus brain factor-1
(MOUSE);, mRNA sequence.

```

AI586081
 VERSION AI586081.1 GI:4571978
 EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 REFERENCE 1 (Bases 1 to 398)
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:645952
 This clone was previously sequenced on the 5' end only, this new data is from the 3' end
 High quality sequence stop: 387.
 Location/Qualifiers
 1..398
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1209608"
 /sex="female"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene mouse skin (#937313)"
 /notes="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
 Alignment Scores:
 Pred. No.: 1..15 Length: 398
 Score: 563.50 Matches: 111
 Percent Similarity: 87.02% Conservative: 3
 Best Local Similarity: 84.73% Mismatches: 16
 Query Match: 26.54% Indels: 1
 DB: 9 Gaps: 1
 US-10-087-080-32 (1-402) x AI586081 (1-398)
 QY 162 TrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValIysValLeuArg 181
 Db 2 TGGCGGCACTCCGTGGCGCACACCTCTCGCTACAGACTGTTCTCAAGGTGCTGCGC 61
 QY 182 AspProSerArgProTrpGlyIleAspAsnTyrTrpMetLeuAsnProAsnSerGluTyr 201
 Db 62 GACCCCTCGCGGCGCTGGGCGCAAGGCAACTACTGATGCTCAACCCCAACGCGAATAC 121
 QY 202 ThrPheAlaAspGlyValPheArgArgArgGlyValArgLeuSerHisArgAlaProVal 221
 Db 122 ACTTTCGCCGACGGGGTCTTCGCCCGCGCGGAGAGCGCTCAGCCACCGGACCCAGATC 181
 QY 222 ProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla---AlaProProPro 240
 Db 182 TCGCGGTCCGGGTGCGGCGCGGAGGAAGGCCACCGGACCTCGCGGAGCCCCCGGAGCCCC 241

```

Qy 241 AlaProAlaAlaProAlaSerProAlaArgMetArgSerProAlaArgGlnGluGluArgAla 260
Db 242 GCGCCGCGCGCGCTCTCCCGATCGCGGCTCGCGGCTCGCGAGGAGCGCTCC 301
Qy 261 SerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
Db 302 AGCCCTCGAGCAGAGTCTCCAGCTCTCGCCATCGACAGCATTCTCAGCAAGCCTTTT 361
Qy 281 ArgSerArgLeuArgAspThrAlaProGly 291
Db 362 CGCAGCGCGCGAGCGGCTCGGCTCGGG 394

RESULT 14
BI443539/c
LOCUS
DEFINITION
dag55b10.x1 Wellcome CRC PRN3 St19 26 egg animal cap Xenopus laevis
cDNA clone IMAGE:4785354 3', similar to SW:HFH1 RAT Q63244
HEPATOCYTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 1-, mRNA sequence.
ACCESSION
BI443539
VERSION
BI443539.1 GI:15268246
KEYWORDS
EST.
ORGANISM
Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 688)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L.,
Pape, D., Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y.,
Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Other ESTs: dag55b10.y1
CONTACT: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, E. ellefroid, and A.M. Zorn
(Wellcome/CRC Institute). DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 417.
FEATURES
Location/Qualifiers
1..688
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4785354"
/tissue_type="egg, subtracted by stage 19-26 animal cap"
/lab_host="DH10B (phage-resistant)"
/clone_vector="Wellcome CRC PRN3 St19 26 egg animal cap"
/notes="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute). "
ORIGIN
Alignment Scores:
Pred. NO.: 2.69e-15 Length: 688
Score: 560.50 Matches: 116
Percent Similarity: 61.14% Conservative: 13
Best Local Similarity: 54.98% Mismatches: 25
Query Match: 26.40% Indels: 57
DB: 12 Gaps: 3
US-10-087-080-32 (1-402) x BI443539 (1-688)

```

```

Qy 117 ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 136
Db 688 CCAAAACTTCTTATTCGATTCATTAATAGCCATGGCTACGAGGACTCGGCCAGT 629
Qy 137 GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPhePheArg 156
Db 628 GCGCCCTGACACTCGGCGAGATCAATGACTATTCATGAAGAATATCCCTTCTTCAGG 569
Qy 157 GlySerTyrThrGlyTTPArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe 176
Db 568 GGTAGCTACACTGGCTGGAGGAACCTCAGTGGGACACATCTCTCCCTCAATGACTGTTT 509
Qy 177 ValLysValLeuArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsn 196
Db 508 GTCAAGTTCACGGGACCCATCTAGGCCATGGGCAAGATATTAATCTGATGCTGAAC 449
Qy 197 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgAArgLysArgLeuSer 216
Db 448 CCAACACGTGAGTACACCTTTGCAGATGGGTCTTCAGACGACAGAGGAGGACTAAAC 389
Qy 217 HisArgAlaProAlaProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAla 236
Db 388 AGGTCCACCAAGTGC-----CTCAAGACGAGGATCTGCAGGACTT----- 347
Qy 237 AlaProProAlaProAlaProAlaSerProArgMetArgSerProAlaArgGln 256
Db 346 -----GCAGAACAGCAGCAGCACAATGATGAACCCAACTAAAGCA 308
Qy 257 GluGluArgAlaSerProAlaGly----- 264
Db 307 AAACAAAGCTGTTCCTCCCTCATCCAGCGCGTGTATCATGGCACATATAAGACATTCATCC 248
Qy 264 ----- 264
Db 247 TCCTCTACAACTCTCTCTCCACAGGCTACCAAGAAACCAACTCGGGGACCAAGTTC 188
Qy 265 -----LysPheSerSerSerPheAlaTle 272
Db 187 TCCAGCAGGCTTACCAAGAAACCAACTCGGGGACCAAGTTCCTCAGCTCTTCGCAATT 128
Qy 273 AspSerIleLeuArgLysProPheArgSerArg 283
Db 127 GAAGATCTTCTGATTAACCTTTCAGAGGAGA 95

RESULT 15
CD282719 542 bp mRNA linear EST 23-MAY-2003
LOCUS G38935.47 NCI_CGAP_Zemb2 Danio rerio cDNA clone IMAGE:6519563 5',
DEFINITION mRNA sequence.
ACCESSION CD282719
VERSION CD282719.1 GI:31060495
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 542)
Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
Expressed sequence tags from NCI_CGAP_Zemb2, a Danio rerio
embryonic library
Unpublished (2003)
Contact: Chen F.
EXELIXIS, Inc.
170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA Sequencing by: Exelixis, Inc. Clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 14103 row: N column: 11

```

Job time : 2587 secs

FEATURES High quality sequence stop: 542.
Location/Qualifiers
source 1..542
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6519563"
/tissue_type="embryo"
/lab_host="DH10B (TI-resistant)"
/clone_lib="NCI CGAP ZEmb2"
/notes="Vector: PCMV-SpOFT6.1; Site 1: EcoRV; Site 2: NotI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2 kb. Constructed by J. Wang (Research Genetics,
Invitrogen Corp) from tissue donated by L. Zon (Harvard
University). Note: this is a NCI CGAP Library."

ORIGIN

Alignment Scores:	1.29e-13	Length:	542
Pred. No.:	519.00	Matches:	112
Score:	62.19%	Conservative:	13
Best Local Similarity:	55.72%	Mismatches:	24
Query Match:	24.45%	Indels:	52
DB:	14	Gaps:	5

US-10-087-080-32 (1-402) x CD282719 (1-542)

```
QY 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
Db ATGAAGTTGGAGGTTTC-----TCTGGAGTCGGCTTGTGGACAAGCGCGTGGACCTG 149
QY 21 GluGlyAlaGlyCysSerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
Db TGCAGTGACATGGATCGCAACGTCGGCTTCCCGCTCTCCACG-----GAGGAGGAATTG 203
QY 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyAlaAlaArgAspThrGln 60
Db GGCTCGGATGGAGACTGCTCAGCAAC-----230
QY 61 GlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaLeProAlaAla 80
Db -----AGTCAGGACCGCGTGCCCT-----251
QY 81 AlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAlaGly 100
Db -----GTTCCGGACGG-----263
QY 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
Db -----AAGCGGAACCATATACGCTAGACCCAAACCACT 299
QY 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
Db TACTCTTACATTCATTTATGCAATGGCCATACGGGATTCCAAACACCGTCGTCTTACA 359
QY 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
Db CTTGCTGAATAAACGATACITTAATGAAAAATTCCTGTTTTCGGGGCAGCTACACG 419
QY 161 GlyTyrArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
Db GGGTGGCGAAACCTCTACGGCATACTTGTCATTAAACGACTGCTTTTAAAGTGTG 479
QY 181 ArgAspProSerArgProTyrGlyLysAspAsnTyrTyrMetLeuAsnProAsnSerGlu 200
Db CGAGATCCCTCGGTCGTCGGGCAAGACCATATTGATGTTGNACCCACACAGCGAG 539
QY 201 Tyr 201
Db 540 TAC 542
```

Search completed: March 31, 2004, 08:38:08

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 30, 2004, 22:00:37 ; Search time 4678 Seconds
(without alignments)
11201.738 Million cell updates/sec
Title: US-10-087-080-31
Perfect score: 1209
Sequence: 1 atgaagcttgaggtgttgt.....tggagacgtcttagcttga 1209

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl: 1: gb_ba*
2: gb_ba*
3: gb_in*
4: gb_om*
5: gb_ov*
6: gb_pat*
7: gb_ph*
8: gb_pl*
9: gb_pr*
10: gb_ro*
11: gb_sts*
12: gb_sy*
13: gb_un*
14: gb_vi*
15: gb_ba*
16: em_fun*
17: em_hum*
18: em_in*
19: em_mus*
20: em_om*
21: em_or*
22: em_ov*
23: em_pat*
24: em_ph*
25: em_pl*
26: em_ro*
27: em_sts*
28: em_un*
29: em_vi*
30: em_htg_hum*
31: em_htg_inv*
32: em_htg_other*
33: em_htg_mus*
34: em_htg_pln*
35: em_htg_rod*
36: em_htg_mam*
37: em_htg_vrt*
38: em_sy*
39: em_htgo_hum*
40: em_htgo_mus*
41: em_htgo_other*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1208	99.9	1448	9	AF153341 Homo sapi
2	1183.8	97.9	62761	9	AL499606 Human DNA
3	1180.6	97.7	2322	9	BC053850 Homo sapi
4	1180.6	97.7	4109	9	AF225950 Homo sapi
5	781.6	64.6	1513	10	AF154426 Mus muscu
6	781.6	64.6	2046	10	BC047155 Mus muscu
7	781.6	64.6	187709	10	AL589738 Mouse DNA
8	780	64.5	4763	10	AF010405
9	773.4	64.0	1576	10	AF153193 Rattus ra
10	773.4	64.0	195184	2	AC119497 Rattus ra
11	768.4	63.6	182152	2	AC067929
12	717.8	59.4	1760	10	RATHFH1
13	653	54.0	178168	2	AC012350 Homo sapi
14	545.4	45.1	74969	2	AC016269 Homo sapi
15	250.8	20.7	182152	2	AC067929
16	227.8	18.8	2026	5	GGU47276 Gallus gall
17	211.8	17.5	182835	10	AC121839 Mus muscu
18	209.4	17.3	173073	9	AC098522 Homo sapi
19	207	17.1	2437	10	MUSBR2F
20	201.4	16.7	2272	9	HSU59832
21	201.4	16.7	5181	6	AX336763 Sequence
22	201.4	16.7	5181	9	HSU59831 Human trans
23	199.6	16.5	2225	10	MMU41047 Mus muscu
24	199.6	16.5	2305	10	AF067421
25	197.8	16.4	2807	10	RNTTF2PRT
26	196.8	16.3	62848	5	BX510653 Zebrafish
27	195.2	16.1	160796	9	AL499604 Human DNA
28	195	16.1	1446	9	AF197560 Homo sapi
29	195	16.1	142728	9	HSU59831 Human trans
30	194.6	16.1	3510	9	HSU59831 Human trans
31	193.4	16.0	22741	2	AC127771 Rattus no
32	193.4	16.0	286524	2	AC105802 Rattus no
33	193	16.0	2011	9	AF275722 Homo sapi
34	193	16.0	106239	9	AL607122 Human DNA
35	192.8	15.9	186734	10	BX005053 Mouse sapien
36	192	15.9	1673	9	HSTIF2
37	189.6	15.7	230905	2	AC132180 Rattus no
38	188.8	15.6	2606	10	MMU457047 Mus muscu
39	188.8	15.6	217762	10	AL806523 Mouse DNA
40	188.6	15.6	232988	2	AC095712 Rattus no
41	188	15.6	4259	9	AF042832 Homo sapi
42	187.8	15.5	153604	2	AC022754 Homo sapi
43	187.6	15.5	1860	6	AR083460 Sequence
44	186.6	15.4	185539	2	AC018349 Homo sapi
45	184.4	15.3	1965	6	AR274884 Sequence

ALIGNMENTS

RESULT 1
AF153341
LOCUS Homo sapiens winged helix/forkhead transcription factor (HFHL)
DEFINITION Homo sapiens winged helix/forkhead transcription factor (HFHL)
gene, complete cds.
ACCESSION AF153341
VERSION AF153341.1
KEYWORDS GI:8489092
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1448)
AUTHORS Hong H.K., Noveroske J.K., Headon D.J., Liu T., Sy M.S.,
Justice M.O. and Chakravarti A.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 22:00:37 ; Search time 4678 Seconds
(without alignments)
11201.738 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaattgagtggttgtt.....tggagacgtcttagttga 1209

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_btg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sta.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1308	99.9	1448	9	AF153341	AF153341 Homo sapi
2	1183.8	97.9	62761	9	AL499606	AL499606 Homo sapi
3	1180.6	97.7	2322	9	BC053850	BC053850 Homo sapi
4	1180.6	97.7	4109	9	AF225950	AF225950 Homo sapi
5	781.6	64.6	1513	10	AF154426	AF154426 Mus muscu
6	781.6	64.6	2046	10	BC047155	BC047155 Mus muscu
7	781.6	64.6	187709	10	AL585738	AL585738 Mouse DNA
8	773.4	64.0	4763	10	AF010405	AF010405 Mus muscu
9	773.4	64.0	1576	10	AF153193	AF153193 Rattus ra
10	768.4	63.6	182152	2	AC119497	AC119497 Rattus no
11	768.4	63.6	182152	2	AC067929	AC067929 Homo sapi
12	717.8	59.4	1760	10	BATPH1	.L13201 Rattus norv
13	653	54.0	178168	2	AC012350	AC012350 Homo sapi
14	545.4	45.1	74969	2	AC016269	AC016269 Homo sapi
15	250.8	20.7	182152	2	AC067929	AC067929 Homo sapi
16	227.8	18.8	2026	5	GGU47276	U47276 Gallus gall
17	211.8	17.5	183835	10	AC121839	AC121839 Mus muscu
18	209.4	17.3	173073	9	AC099522	AC099522 Homo sapi
19	207	17.1	2437	10	MUSBE2TF	L38602 Mus musculu
20	201.4	16.7	2272	9	HSU59832	U59832 Human trans
21	201.4	16.7	5181	6	AX336763	AX336763 Sequence
22	201.4	16.7	5181	9	HSU59831	U59831 Human trans
23	199.6	16.5	2225	10	MMU41047	U41047 Mus musculu
24	199.6	16.5	2305	10	AF067421	AF067421 Mus muscu
25	197.8	16.4	2807	10	RNTT2PRT	Y11321 R.norvegicu
26	196.8	16.3	62848	5	BX510653	BX510653 Zebrafish
27	195.2	16.1	160796	9	AL499604	AL499604 Human DNA
28	195	16.1	1446	9	AF197560	AF197560 Homo sapi
29	195	16.1	142728	9	HSU89995	U89995 Human DNA
30	194.6	16.1	3510	9	HSU89995	U89995 Human DNA b
31	193.4	16.0	227341	2	AC127771	AC127771 Rattus no
32	193.4	16.0	286524	2	AC105802	AC105802 Rattus no
33	193	16.0	2011	9	AF275722	AF275722 Homo sapi
34	193	16.0	108239	9	AL607122	AL607122 Human DNA
35	192.8	15.9	186734	10	BX005053	BX005053 Mouse DNA
36	192	15.9	1673	9	HSTIF2	Y13386 Homo sapien
37	189.6	15.7	230905	2	AC132180	AC132180 Rattus no
38	188.8	15.6	2606	10	MMU457047	AJ457047 Mus muscu
39	188.8	15.6	217762	10	AL806523	AL806523 Mouse DNA
40	188.6	15.6	232988	2	AC095712	AC095712 Rattus no
41	188	15.6	4259	9	AF042832	AF042832 Homo sapi
42	187.8	15.5	153604	2	AC022754	AC022754 Homo sapi
43	187.6	15.5	1860	6	AR083460	AR083460 Sequence
44	186.6	15.4	185539	2	AC018349	AC018349 Homo sapi
45	184.4	15.3	1965	6	AR274884	AR274884 Sequence

ALIGNMENTS

RESULT 1
AF153341
LOCUS AF153341
DEFINITION Homo sapiens winged helix/forkhead transcription factor (HFX1)
ACCESSION AF153341
VERSION AF153341.1
KEYWORDS GI:8489092
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1448)
AUTHORS Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., SY,M.S., Justice,M.J. and Chakravarti,A.

TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice
 JOURNAL Genesis 29 (4), 163-171 (2001)
 MEDLINE 21207067
 PUBMED 11309849
 REFERENCE 2 (bases 1 to 1448)
 AUTHORS Hong, H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA

FEATURES

Location/Qualifiers
 1..1448
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p25"
 <233..51441
 /gene="HFH1"
 <233..>1441
 /gene="HFH1"
 /product="winged helix/forkhead transcription factor"
 233..1441
 /gene="HFH1"
 /codon_start=1
 /product="winged helix/forkhead transcription factor"
 /protein_id="AAF75586.1"
 /db_xref="GI:8489093"
 /translation="MKLEVFVRAAHGDKQSDLEGAGSDAPSLSAAGDSLSGQDCAAKPSAGGARDTQGGEGSAGGPGAEPAIAAAVVAEGRAAGPAGGA GSGEARSKPYTRRKPYSYIALAMAIRSAGGRUTLAENIYLMGKFFFRGSYT GWRNSVRHNSLUNDFVKVLRDPSPWGNKDNMLNPNSEYTFADGPFRRRKLKLSHR APVPAPGRUEEAPGLPAAPPAPAPASPRMRSPARQEEASPAQKFSFSAIDSL RXPFRSLRUDTAPTTQLQWGAAPCPPLPAFPALLPAAPCPALLPLCAYGAGEPARLG AREAEVPTAPTPPLLALPLPAAPAKPLRGAAGHAHYCLRLPLAALQAALVRRPGH LSPVETILLA"

ORIGIN

Query Match 99.9%; Score 1208; DB 9; Length 1448;
 Best Local Similarity 100.0%; Pred. No. 1.1e-120;
 Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTTGGAGTGTTCCTCCTCGCGCGGCCACGGGACAAAGCAGGCGAGTACCTG 60
 DB 233 ATGAAGTTGGAGTGTTCCTCCTCGCGCGGCCACGGGACAAAGCAGGCGAGTACCTG 292

QY 61 GAGGCGCGCGGCGGAGAGACGCGCGTCCCGCTGTTCGGCGGCGGAGACGACTCCCTG 120
 DB 293 GAGGCGCGGCGGCGGAGAGACGCGCGTCCCGCTGTTCGGCGGCGGAGACGACTCCCTG 352

QY 121 GGCTCAGATGGGAGCTGCGCGGCCAAGCGTCCGCGGCGGCGGCGGCGGAGATACGAG 180
 DB 353 GGCTCAGATGGGAGCTGCGCGGCCAAGCGTCCGCGGCGGCGGCGGCGGAGATACGAG 412

QY 181 GGCAGCGGCGAACAAGATGTCGGAGGCGCGGCGGCGGCGGAGGCGGATCCCGGACGA 240
 DB 413 GGCAGCGGCGAACAAGATGTCGGAGGCGCGGCGGCGGCGGAGGCGGATCCCGGACGA 472

QY 241 GTGTGTGACGCGGTGTCGGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 DB 473 GTGTGTGACGCGGTGTCGGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 532

QY 301 GCGCGGGGAGCGGCGGAGGTCACGCGAGCAGCCATATAGCGGCGGCGGCCCAAGCCCC 360
 DB 533 GCGCGGGGAGCGGCGGAGGTCACGCGAGCAGCCATATACGCGGCGGCGGCCCAAGCCCC 592

QY 361 TACTGTATACGCGGTCTATCCCATGCGCATCCCGCATCCGCGGCGGCGGCGGCGGTACG 420
 DB 593 TACTGTATACGCGGTCTATCCCATGCGCATCCCGCATCCGCGGCGGCGGCGGCGGTACG 652

QY 421 CTGGCGGAGATCAACGAGTACCTCTAGGCGAGTTCCTCTTTTCCGCGGCGAGCTACAG 480
 DB 653 CTGGCGGAGATCAACGAGTACCTCTAGGCGAGTTCCTCTTTTCCGCGGCGAGCTACAG 712

QY 481 GGCTGGCGCAACTCCGTTGGCGCCAAACCTTTGCTCAACGACTGCTTCTGTCFCAAGGTGCTG 540
 DB 713 GGCTGGCGCAACTCCGTTGGCGCCAAACCTTTGCTCAACGACTGCTTCTGTCFCAAGGTGCTG 772

QY 541 CGCGACCCCTCGCGGCGCTGGGCGCAAGGACAACTACTGGATGCTCAACCCCAACAGCGAG 600
 DB 773 CGCGACCCCTCGCGGCGCTGGGCGCAAGGACAACTACTGGATGCTCAACCCCAACAGCGAG 832

QY 601 TACACTTTCGCCGACGGGGTCTTCGCGCGCGCGCAAGCGCTCTCAGCCACACGCGGCGCG 660
 DB 833 TACACTTTCGCCGACGGGGTCTTCGCGCGCGCGCAAGCGCTCTCAGCCACACGCGGCGCG 892

QY 661 GTCCCGCGCGCGCGGCTGGCGCGCGGAGGAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCC 720
 DB 893 GTCCCGCGCGCGCGGCTGGCGCGCGGAGGAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCC 952

QY 721 GCGCGCGCGCGCGCGCTCGCGCGCGCATGCGCTCGCGCGCGCGCGCGCGAGGAGCGCGCC 780
 DB 953 GCGCGCGCGCGCGCGCTCGCGCGCGCATGCGCTCGCGCGCGCGCGCGCGAGGAGCGCGCC 1012

QY 781 AGCCCGCGCGGCAAGTCTCCAGCTCTTCGCGCATCGACAGCATCTTGGCGAAGCCCTTC 840
 DB 1013 AGCCCGCGCGGCAAGTCTCCAGCTCTTCGCGCATCGACAGCATCTTGGCGAAGCCCTTC 1072

QY 841 CGCAGCGCTCGCTCAGGGACACGGCGCGCGCGCGCGCTTCAGTGGGCGCGCGCGCGCC 900
 DB 1073 CGCAGCGCTCGCTCAGGGACACGGCGCGCGCGCGCGCTTCAGTGGGCGCGCGCGCGCC 1132

QY 901 TGC CGCGCGCTGCGCGCGTTCGCGCGCTCTCCCGCGCGCGCTTCGCGCGCGCTTCGCTG 960
 DB 1133 TGC CGCGCGCTGCGCGCGTTCGCGCGCTCTCCCGCGCGCGCTTCGCGCGCGCTTCGCTG 1192

QY 961 CCGCTCTGCGCGTACGCGCGCGCGCGCGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGG 1020
 DB 1193 CCGCTCTGCGCGTACGG 1252

QY 1021 CCACCGACCGCGCGCGCTCTCTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCA 1080
 DB 1253 CCACCGACCGCGCGCGCTCTCTGCTTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCA 1312

QY 1081 CTCGAGGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCG 1140
 DB 1313 CTCGAGGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCG 1372

QY 1141 CTCGAGGCGCGCTTACTGCGNCGTCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 1200
 DB 1373 CTCGAGGCGCGCTTACTGCGNCGTCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 1432

QY 1201 CTAGCTTGA 1209
 DB 1433 CTAGCTTGA 1441

RESULT 2

AL499606 62761 bp DNA linear PRI 19-DEC-2001
 LOCUS Human DNA sequence from clone Rp11-13J16 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL499606
 VERSION AL499606.18 GI:17973944
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 62761)
 Corby, N.
 Direct Submission
 Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Dec 20, 2001 this sequence version replaced gi:17902904.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-13U16 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-13U16 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-13U16 is at 1 in this sequence. The true left end of clone RP4-668J24 is at 60762 in this sequence. The true right end of clone RP1-116B8 is at 4171 in this sequence.
Location/Qualifiers
1..62761
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-13U16"
/clone_lib="RP11-13U16"
17171..17234
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
17634..17641
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
21165..21540
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Assembly consistent with restriction digest data."
21466
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 430bp by restriction digest data."
23824..24045
/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
60092
/note="Tandem repeat. Forced join. Gap size estimated to be approximately 120bp by restriction digest data."
ORIGIN
Query Match 97.9%; Score 1183.8; DB 9; Length 62761;
Best Local Similarity 99.1%; Pred. No. 1.2e-118;
Matches 1201; Conservative 0; Mismatches 8; Indels 3; Gaps 1;
QY 1 ATGAAGTTGGAGGTGTTCTTCCTCGCGGCCACGGGACACAGGAGGACGAGGACCTG 60
Db 16945 ATGAAGTTGGAGGTGTTCTTCCTCGCGGCCACGGGACACAGGAGGACGAGGACCTG 17000
QY 61 GAGGCGCGCGCGGACGACGACGCGCGCTCCCTCGCGCGGCGGAGACGACTCCCTG 120

Db	18085	GCCTTCAGCGCGCTCAGTTCGCGCCCTCGCCGACCTCCCTGAGACG	18144
QY	1198	CTCCTGACTGTGA	1209
Db	18145	CTCCTGAGCTGA	18156
RESULT 3			
LOCUS	BC053850		
DEFINITION	Homo sapiens forkhead box Q1, mRNA (cDNA clone MGC:61699 IMAGE:6010412), complete cds.		
ACCESSION	BC053850	2322 bp	mRNA linear PRI 11-DEC-2003
VERSION	BC053850.1	GI:31753144	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2322)		
	Strausberg,R.L., Peingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shemmen,C.M., Schuler,G.D., Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldi,M.P., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toehiyuki,S., Carrinzi,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickinson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywicki,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.		
TITLE	Generation and initial analyses of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903	(2002)
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2322)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTF/Gazdar cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@nhgri.nih.gov Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Grante,S., Guan,X., Gupta,J., Hagnighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantiripop,S., Thomas,P.J., Touchman,J.W., Tsourgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.		
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAP Plate: 115 Row: 1 Column: 14 This clone was selected for full length sequencing because it			

QY 478 ACGGCTGGCGCAACTCCGTGGCGCCACAACTTTCGCTCAACGACTGCTTCGTCAAGGTG 537
Db 2217 ACGGCTGGCGCAACTCCGTGGCGCCACAACTTTCGCTCAACGACTGCTTCGTCAAGGTG 2276
QY 538 CTGGCGCAACCTCCGTGGCGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGC 597
Db 2277 CTGGCGCAACCTCCGTGGCGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGC 2336
QY 598 GAGTACACCTTCGCGGACGCGGCTTCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGG 657
Db 2337 GAGTACACCTTCGCGGACGCGGCTTCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGG 2396
QY 658 CCGTTCGCGCGCGCGCGCTGCGCGCGCGAGAGCGCGCGCGCTCCCGCGCGCGCGCG 717
Db 2397 CCGTTCGCGCGCGCGCGCTGCGCGCGCGAGAGCGCGCGCGCTCCCGCGCGCGCGCG 2456
QY 718 CCGCGCGCGCGCGCGCTGCGCGCGCGATGCGCTGCGCGCGCGCGCGCGCGCGCGCG 777
Db 2457 CCGCGCGCGCGCGCGCTGCGCGCGCGATGCGCTGCGCGCGCGCGCGCGCGCGCGCG 2516
QY 778 GCGAGCGCGCGCGCGCTTCAGCTCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 837
Db 2517 GCGAGCGCGCGCGCGCTTCAGCTCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2576
QY 838 TTCCGCGCGCGCGCGCTCAGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897
Db 2577 TTCCGCGCGCGCGCGCTCAGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2636
QY 898 CCTGCGCGCGCGCGCTGCGCGCGCGTCCCGCGCGCTCCCGCGCGCGCGCGCGCGCGCG 957
Db 2637 CCTGCGCGCGCGCGCTGCGCGCGCGTCCCGCGCGCTCCCGCGCGCGCGCGCGCGCGCG 2696
QY 958 CTGCGCGCTGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1017
Db 2697 CTGCGCGCTGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 2756
QY 1018 GTGCCACGACCGCGCGCGCTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1077
Db 2757 GTGCCACGACCGCGCGCGCTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2816
QY 1078 CCACTCCGAGCG 1137
Db 2817 CCACTCCGAGCG 2876
QY 1138 GCGTCCGAGCGCGCGCTTAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197
Db 2877 GCGTCCGAGCGCGCGCTTAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2936
QY 1198 CTCCTAGCTTGA 1209
Db 2937 CTCCTAGCTTGA 2948

RESULT 5
AF154426 1513 bp DNA linear ROD 24-AUG-2001
LOCUS AF154426
DEFINITION Mus musculus HFH1 (Hfh1) gene, partial cds.
ACCESSION AF154426
VERSION AF154426.1 GI:8347144
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1513)
AUTHORS Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
Justice,M.J. and Chakravarti,A.
TITLE The winged helix/forkhead transcription factor Foxq1 regulates
differentiation of hair in satin mice
JOURNAL Genes; 29 (4), 163-171 (2001)
MEDLINE 21207067
PUBMED 11309849
REFERENCE 2 (bases 1 to 1513)

AUTHORS Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
TITLE Direct Submission
JOURNAL Submitted (26-MAY-1999) Genetics, Case Western Reserve University,
10900 Euclid Avenue, Cleveland, OH 44106, USA
FEATURES
Location/Qualifiers
1..1513
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
/db_xref="taxon:10090"
<158..>1360
/gene="Hfh1"
<158..>1360
/gene="Hfh1"
/product="Hfh1"
158..1360
/gene="Hfh1"
/codon_start=1
/product="Hfh1"
/protein_id="AAF74524.1"
/db_xref="GI:8347145"
/translation="MKLEVFVPRAAHGDVKMSDLEGAGSSDVPSPLAAGDDSKGSDG
DCAANSPAGAGDLEGGGSSGPPSAQDGPFAIDRRTIOASGACGAGVGG
EGASKPYTRKPPYIYALIAIRDSAGRLTAEINIEYLMKGFPPFGSTYGM
NSVRHLSLNDKLVLDSPKWNMLNPNSEYTFADGFRFRRLKSLRITV
SASGLPEEAPPGAGTQPAAPARSPARQERSSPASKFSSFAIDSLK
PFRSRDGSALGVLPWGAAPCPPLRAYPALLPAAFGALLPLCAVGAPEPTLLASR
GTEVQPAAPALLAFLSTAAPKPRGPTAGAAHLYCLRLPTALQAAAAACGPGPHLS
YPVETLLA"
ORIGIN
Query Match 64.6%; Score 781.6; DB 10; Length 1513;
Best Local Similarity 81.2%; Pred. No. 4.9e-75;
Matches 991; Conservative 0; Mismatches 200; Indels 30; Gaps 6;
QY 1 ATGAATGTGGAGGTTCGTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 158 ATGAATGTGGAGGTTCGTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 217
QY 61 GAGGCG 120
Db 218 GAGGCG 277
QY 121 GCTCAGATGGGAGTGGCG 177
Db 278 GCTCAGATGGGAGTGGCG 337
QY 178 CA---GGCGGACGCGGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 234
Db 338 GAAGTGGCG 397
QY 235 GCAGCAGTGTGCGAGCGGTGGCG 294
Db 398 GCACTGATG-----ACAGCAGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
QY 295 GCGGCG 354
Db 443 GCGGCG 502
QY 355 CCCCCCTACTCGTACATCGCGCTCATCGCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
Db 503 CCCCCCTACTCGTACATCGCGCTCATCGCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 562
QY 415 TTGACGCTGGCGGATCAACGAGTACTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474
Db 563 CTGACACTGGCGGATCAACGAGTACTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 622
QY 475 TACAGCGGCTGGCGCAACTCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
Db 623 TACAGCGGCTGGCGCAACTCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 682
QY 535 GTGCTGCG 594


```
/chromosome="13"  
/clone="RP23-322J11"  
/clone_lib="RPCI-23"
```

ORIGIN

Query Match	64.6%	Score 781.6	DB 10	Length 187709	
Best Local Similarity	81.2%	Pred. No. 9.4e-76			
Matches 991	Conservative	0	Mismatches 200	Indels 30	Gaps 6
QY	1	ATGAAGTTGGAGGTGTTTCGTCCCTCGCGCGGCCCAAGGGGACAAGGAGGCGCAGTGACCTG	60		
DB	135706	ATGAATTTGGAGGTGTTTCGTCCCAACGCGACGCCACGGGACAAATGGCAGCGATCTG	135765		
QY	61	GAGGCGCGGGGGAGCGACGCGCGCTCCCGCTGTTCGGCGCGGGAGACGACATCCCTG	120		
DB	135766	GAGGGGGCGGACGACGAGCGTGCCTATCTCACTGTCTCGCGCTGTGACGACTCTCTTA	135825		
QY	121	GCGTCAGATGCGGACTGCGCGGCCAA---GCCCTCGCGCGCGCGCGGCCCAAGATACG	177		
DB	135826	GACTCAGACGGGACTGTGTGACGCCAAACAGCCGCGCGCGGACGCGCGCGGGATCTG	135885		
QY	178	CA---GGCGAGCGGGAACAGAGTCGGGAGCGGGCCGCGCGGAGGAGCGGATCCCG	234		
DB	135886	GAGGTGCGCGCGGAGAGAAATTCAGTGTGGCGGCGGAGCGCCCAACACGGTTCGGAG	135945		
QY	235	GACGACAGCTGTGACGCGGTGTCGGGAGCGGCGGAGCGCGGGCGCGCGGGCCAGCG	294		
DB	135946	GCAACTGATG-----ACAGCAGAACGCAGAGCCTCCCGCGACGGGCGGTGC	135990		
QY	295	GCGGCGCGCGGGAGCGGCGAGGTGCAACGACGAGCCATATACGGGGGGGCCCAAG	354		
DB	135991	GCGGCGCGGTGGGCGGCGCGGAGGGCGCGGACGAGCGGTACACGGCGGGCGCCCAAG	136050		
QY	355	CCCCCTACTCTGATCGCGCTATCGCCATGCGCATCCGCGACTCGCGCGCGGGCGGC	414		
DB	136051	CCCCCACTCTCTCATCGCTCTCATCGCATGCCATCCGCGACTCCGCGCGCGGACGC	136110		
QY	415	TTGACGCTGGCGGAGATCAACGAGTACTCATGGGCAAGTTCCTCTTTTCGCGCGGACG	474		
DB	136111	CTGACACTGGCGGAGATCAACGAGTACTCATGGGCAAGTTCCTCTTTTCGCGGGCAGC	136170		
QY	475	TACCGGGTGGCGCAACTTCGGTGGCGGCACAACTTTTCGCTCAACGACTGCTTCGTCAAG	534		
DB	136171	TACACGGGTGGCGCAACTTCGGTGGCGGCACAACTTCGCTCAACGACTGTTTCGTCAAG	136230		
QY	535	GTGCTGGCGCAACCTTCGGGCGCTTGGGGCAAGGACAATCTACTGATGCTCAACCCCAAC	594		
DB	136231	GTGCTGGCGGACCTTCGCGGCGCTTGGGGCAAGGACAATCTACTGATGCTCAACCCCAAC	136290		
QY	595	AGCGAGTACACTTCGCGACGGGTCTTCGCGCGCGCGCGCAAGGCGCTTCAGCCACCGC	654		
DB	136291	AGCGAATACTTCGCGACGGGTCTTCGCGCGCGCGCGCAAGGCGCTTCAGCCACCGC	136350		
QY	655	GGCGCGTCCCGCGCGCGGCTGGCGCGCGGAGGCGCGCGGCTCCCGCGCG---CG	711		
DB	136351	ACCACAGTCTCGCGGTTCGGGCGTGGCGCGGAGGAGCCCAACCGGACCTTCGCGGACCC	136410		
QY	712	CGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCATGGCTCGCGCGCGCGCGCGCGCGAG	771		
DB	136411	CGCGAGCGCGCGCGCGCGCGCGCTTCCTCCCGATCGCGCGCTCGCGCGGCTCGCCAGGAG	136470		
QY	772	GAGCGCGCAGCGCGCGGGCAAGTTCATCGACTCTTCGCGCATCGACAGCATCTTCGCGC	831		
DB	136471	GAGCGCTCCAGCGCTTCGAGCAAGTTCATCGACTCTTCGCGCATCGACAGCATCTTCAGC	136530		
QY	832	AAGCCCTTCGCGACGCGTCCGCTCAGGAGACACGGGCCCGGGGACGAGCTTCAGTGGGGC	891		
DB	136531	AAGCCTTTCGCGACGCGCGCGAGCGGACCTCGGCTCTGGGGTGCAGCTACCTCGGGC	136590		
QY	892	GCGCGCGCTTCGCGCGCTTCGCGGTTCCCGGCTCTCTCCCGCGGGCGCCCTCAGG	951		
DB	136591	GCGGCTCTCTTCGCGCGCTTCGCGGCTATCCCGGCTCTCTTCGCGCGGGCGCGGTGCG	136650		

Qy	952	GCCTGCTGCCGTCTCCGGCGTACGGCGCGGAGCGCGCGCTGGGGCGCGCGAG	1011
Dd	136851	GCTCTGCTACCGTCTGTGTCTTACGGCGCAAGCGAGCCTACGCTGCTGGCTCGCGCGG	136710
Qy	1012	GCGAGGTGCACCGACCGCGCGCCCCCTCTGCTTGTCACCTCTCCGCGCGCGCCCC	1071
Dd	136711	ACGAGGTG---CAGCCCGCGCGCCCCTTCTGCTGCGCGCCCTCTCCACCGCGCTCCA	136767
Qy	1072	GCCAAGCCATTCGAGAGCCCGG-----CGGCGCGCGCGCGCACCTGTACTGCCCCCTGGG	1128
Dd	136768	GCCAAGCCATTTCGAGGTCCGGAGACCGCGCGCGCGCACCTGTACTGCCCCCTACGG	136827
Qy	1129	CTGCCCCGACGCTTCGAGCGGCGCTTAGTTCGNCGTCTGTCGCGCGCGCACCTGTCTGCTACC	1188
Dd	136828	CTGCCACAGCCCTCGAGCGGAGCGGCGCTGCGGTTCCGCGTCCGCGCACCTGTCTTACC	136887
Qy	1189	GTGGAGACGCTCTTACGTTGA	1209
Dd	136888	GTGGAGACTCTGCTAGCTTGA	136908
RESULT 8	AF010405	4763 bp DNA linear ROD 20-JAN-2000	
LOCUS	AF010405	Mus musculus fork head transcription factor (Hfh-1L) gene, complete cds.	
DEFINITION	AF010405	GI:6718869	
ACCESSION	AF010405	Mus musculus (house mouse)	
VERSION	AF010405.2	GI:6718869	
KEYWORDS		Mus musculus	
SOURCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;	
ORGANISM		1 (bases 1 to 4763)	
REFERENCE		Frank, S. and Zöll, B.	
AUTHORS		Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal location, and expression in adult and embryonic kidney	
TITLE		DNA Cell Biol. 17 (8), 679-688 (1998)	
JOURNAL		98392851	
MEDLINE		9726250	
PubMed		2 (Bases 1 to 4763)	
REFERENCE		Frank, S. and Zöll, B.	
AUTHORS		Direct Submission	
TITLE		Submitted (19-JUN-1997) University of Goettingen, Institute of Human Genetics, Gosselerstr 12d, Goettingen 37073, Germany	
JOURNAL		3 (bases 1 to 4763)	
REFERENCE		Pasche, B., Bieller, A. and Zöll, B.	
AUTHORS		Direct Submission	
TITLE		Submitted (19-JAN-2000) University of Goettingen, Institute of Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany	
JOURNAL		Sequence update by submitter	
REMARK		On Jan 20, 2000 this sequence version replaced gi:2997586.	
COMMENT		Location/Qualifiers	
FEATURES		1..4763	
source		/organism="Mus musculus"	
		/mol_type="genomic DNA"	
		/strain="129"	
		/db_xref="taxon:10090"	
		/chromosome="13"	
		/tissue type="kidney"	
		<2467...>3669	
gene		/gene="Hfh-1L"	
		/note="HNF-3/fork head homologue-1 like"	
		<2467...>3669	
mRNA		/gene="Hfh-1L"	
		/product="fork head transcription factor"	
		2467...3669	
CDS		/gene="Hfh-1L"	
		/note="member of HNF-3/fork head transcription factor family"	
		/codon_start=1	
		/product="fork head transcription factor"	
		/protein_id="AAC12973.2"	


```
/db_xref="GI:6716870"
/translation="MKLEVPAPRAAHGDKMGSDLLEGAGSSDVPSPLSAAGDSDLSGSDG
DCAANSPAGSAGDLEGGGSGSPSADQPEATDSDRTOASAAAGPAGGAGVGG
EGARSKPYRRPKPYVSYTALIAMRDSAGRLTAEINELVKGKFFFRGSGVTGWR
NSVRHLSLNDCEVLEDPSPKWKDNYMLNPSEVTFADGVFRERKLSHRTTV
SASGLRPERAPGAGTPOAAPARSSPIARSPARQBERSPASKESSFALDSLSK
PFRSRDGSALGVQIPWGAACPPLRAYPALLPAAGCALPLCYAGASEPTLLASR
GTEVQPAAPLLAPLAPLSTAAAPKFRFPETAGAAHLYCFLRLPTALQAAAACGPGFHL
YPVETLLA"
misc_feature
2782..3123
/genes="Hfh-1L"
/note="encodes DNA binding domain"

ORIGIN
Query Match 64.5%; Score 780; DB 10; Length 4763;
Best Local Similarity 81.1%; Pred. No. 4.9e-75;
Matches 990; Conservative 0; Mismatches 201; Indels 30; Gaps 6;

QY 1 ATCAAGTTGAGGTGTTGTCCTCGCGCGGCCACGGGACAAAGCAGGCGAGTGACCTG 60
Db 2467 ATGAATTTGAGGTGTTGTCCTCGCGCAGCGCCAGCGGAGCAAAATGGCAGCGATCTG 2526
QY 61 GAGGCGCGCGGCGGCGGACGACGCGCGCTCCCGCTGTCGGCGCGGAGACGACTCCCTG 120
Db 2527 GAGGCGCGCGGCGGCGGACGACGCGCTCCCACTGTCGCGCGCTGGTGACGACTCCTTA 2586
QY 121 GGCTCAGATGGGACTCGCGCGGCA--GCCGTCCGCGCGCGCGGCGGCCACAGATACG 177
Db 2587 GGCTCAGACGGGACTGTGCGCGCAACAGCCCGCGCGGCGGAGCGCCCGGATCTG 2646
QY 178 CA---GGGCGACGCGGAACAGAGTCGGGAGCGCGGCGCGGCGGAGGAGCGATCCCG 234
Db 2647 GAAGGTGGCGCGCGGCGGAGGAAATTCGAGTGGCGGCGGCGGCGGCGGCGGCGGCG 2706
QY 235 CGACGAGCTCTCGACGGTGTGCGGAGGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGG 294
Db 2707 GCAACTGATG-----ACAGCAGAACCGACGCGCTCCCGCGGAGCGCGCTG 2751
QY 295 GCGGCGCGCGCGGCGGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 354
Db 2752 GCGGCGCGCGTGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2811
QY 355 CCGCCCTACTGTACATCGCGCTCATCGCATGCGCATCGGACTCGGACTCGGCGGCGGCGG 414
Db 2812 CCGCCCTACTGTACATCGCGCTCATCGCATGCGCATCGGACTCGGCGGCGGCGGCGG 2871
QY 415 TTGACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCCCTTTTTCGCGGCG 474
Db 2872 CTGACACTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCCCTTTTTCGCGG 2931
QY 475 TACAGGGGCTGGCGCAACTCCGTGGCGCACAACTTTCGCTCAAGACTGCTTCGTCAG 534
Db 2932 TACAGGGGCTGGCGCAACTCCGTGGCGCACAACTTTCGCTCAAGACTGCTTCGTCAG 2991
QY 535 GTGCTGGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGGATGCTCAACCCCAAC 594
Db 2992 GTGCTGGCGGACCCCTCGCGGCGCTGGGCGAAGCAACTACTGGATGCTCAACCCCAAC 3051
QY 595 AGCGAGTACACTTTCGCGGAGGGGTTCCTCGCGCGCGCGCGGCGGCGGCGGCGGCGG 654
Db 3052 AGCGAATACACTTTCGCGGAGGGGTTCCTCGCGCGCGCGCGGCGGCGGCGGCGGCGG 3111
QY 655 GCGCGGCTCCCGCGCGCGGCGGCTGGCGCGGAGGAGCGCGGCGGCGGCGGCGGCGG 711
Db 3112 ACCAAGTCTCCGCGTCCGCGGCTGGCGCGGAGAGAGCCCAACCGCGGACTCGCGGAG 3171
QY 712 CCGCGCGCGCGCGCGCGCGCGGCTCGCGCGGATGCGCTCGCGCGCGCGCGCGCGCGG 771
Db 3172 CCGCAGCGCGCGCGCGCGCGCGCGGCTCGCGCGGATGCGCTCGCGCGCTCGCGGAG 3231
QY 772 GAGCGCGCGAGCGCGCGGCGGAGTTCTCCAGTCTCTTCCGATCCAGCAGCATCTCGCG 831
Db 3232 GAGCGCTCCAGCGCGCTGGCGGCAAGTTCTCCAGTCTCTTCCGATCCAGCAGCATCTCAG 3291
```

```
QY 832 AAGCCCTTCGCGAGCGCTCGCCTCAGGACACCGGCCCCCGGAGCGAGCTTTCAGTGGGGC 891
Db 3292 AAGCCTTTTCGCGAGCGCGCGCGAGCGGAGCTCGGCTCTGGGGGTGCGAGTACCTGGGGC 3351
QY 892 GCGCGCGCTTCGCGCGCGCTCGCGGCTTCCCGCGCTCTCCCGCGCGCGCGCTTCAGG 951
Db 3352 GCGCTCTCTTCGCGCGCGCTTCGCGGCTATCCCGCGCTCTCTCCCGCGCGCGCGTGC 3411
QY 952 GCGCTGTGCGCTTCGCGGTACGCGCGCGGCGGAGCGCGCGGCTGGGCGGCGCGGAG 1011
Db 3412 GCTCTGCTACCGCTCTGTGCTTACGCGCGAAGCGAGCTTACGCTGTGGGCTCGCGCGG 3471
QY 1012 GCGGAGGTGCGACCGCGCGCGCTCTCTGCTTGCACCTCTCCCGCGCGGCGCGCCCC 1071
Db 3472 ACCGAGTG---CAGCCCGCGGCGCGCTTCTGCTGGCGCGCTCTCCACCGGCGCTCCA 3528
QY 1072 GCCAAGCACTTCGAGCGCGCG---CGCGCGCGCGCGCGACCTGTACTGCGCGCTCGG 1128
Db 3529 GCCAAGCACTTCGAGGTTCGCGGAGACCGCGCGCGCGCGACCTGTACTGCGCGCTACG 3588
QY 1129 CTGCGCGCGCGCTTCGAGCGCGCTTAGTCCGCGCTCTCGCGCGCGACCTGTCTGTA 1188
Db 3589 CTGCGCGCGCGCTTCGAGCGCGCGCTGCGGCTCCCGGTCGCGGCTGCTTCTTCTAC 3648
QY 1189 GTGAGAGCGCTCTTCTAGCTTGA 1209
Db 3649 GTGAGAGCTCTGCTAGCTTGA 3669

RESULT 9
AF153193 1576 bp DNA linear ROD 24-AUG-2001
LOCUS Rattus rattus winged helix/forkhead transcription factor HFH1
DEFINITION AF153193
ACCESSION AF153193
VERSION AF153193.1 GI:8132282
KEYWORDS Rattus rattus (black rat)
ORGANISM Rattus rattus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Hong,H.-K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
Justice,M.J. and Chakravarti,A.
TITLE The winged helix/forkhead transcription factor Foxq1 regulates
differentiation of hair in satin mice
JOURNAL Genesis 29 (4), 163-171 (2001)
MEDLINE 21207067
PMID 11309849
REFERENCE 2 (bases 1 to 1576)
AUTHORS Hong,H.-K., Noveroske,J.K., Justice,M.J. and Chakravarti,A.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Genetics, Case Western Reserve University,
10900 Euclid Ave., Cleveland, OH 44106-4955
FEATURES
Source
1. 1576
/organism="Rattus rattus"
/mol_type="genomic DNA"
/db_xref="taxon:10117"
<95..>1297
/gene="Hfh1"
<95..>1297
/gene="Hfh1"
/product="winged helix/forkhead transcription factor HFH1"
95..1297
/gene="Hfh1"
/codon_start=1
/product="winged helix/forkhead transcription factor HFH1"
/protein_id="AAF73234.1"
/db_xref="GI:8132283"
/translation="MKLEVPAPRAAHGDKMGSDLLEGAGSSDVPSPLSAAGDSDLSGSDG
```


Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H., Shetty, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Wooley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 195184)
Worley, K.C.

Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195184)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23195197.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWQI
Center clone name: CH230-521F8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 158913 bases at least Q40
Consensus quality: 162109 bases at least Q30
Consensus quality: 164133 bases at least Q20
Estimated insert size: 161675; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 16506: contig of 16506 bp in length
* 16507 16506: gap of unknown length
* 16507 16506: contig of 164032 bp in length
* 180639 180738: gap of unknown length
* 180739 182447: contig of 1709 bp in length

* 182448 182547: gap of unknown length
* 182548 195184: contig of 12637 bp in length.

FEATURES
source
1..195184
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-521F8"
466..1181
/note="clone_boundary"
clone_end:T7
site:
end_sequence:BZ266854"
1768..2639
/note="clone_boundary"
clone_end:Sp6
site:
end_sequence:BZ267695"
16507..17640
/note="wgs_end_extension"
clone_end:Sp6"
71528..72576
/note="wgs_end_extension"
clone_end:Sp6"
76693..78955
/note="wgs_end_extension"
clone_end:Sp6"
86077..87329
/note="wgs_end_extension"
clone_end:Sp6"
106905..108354
/note="wgs_end_extension"
clone_end:Sp6"
129622..130958
/note="wgs_end_extension"
clone_end:Sp6"

ORIGIN
Query Match 64.0%; Score 773.4; DB 2; Length 195184;
Best Local Similarity 80.2%; Pred. No. 7e-75;
Matches 977; Conservative 0; Mismatches 217; Indels 24; Gaps 5;

QY 1 ATGAAGTTGGAGGTTCGTCCTCGCGCGCCACGGGACAGCGGCGGAGTGGGAGCTG 60
Db 136153 ATGAATTTGGAGGTATTTCCTCCACGCGCAGCCACGAGTGGGAGTGGGAGCTG 136212
QY 61 GAGGCGCGCGCGGAGCGCGCGCTCCCGCTGTCGGCGCGGCGGAGAGTCCCTG 120
Db 136213 GAGGCGCGCGGAGCGCGCGCTCCCGCTGTCGGCGCGGCGGAGAGTCCCTG 136272
QY 121 GGTCTAGATGGGACTCGCGCGCCAA---GCCGTCCGGCGCGCGCGCGCGGAGATACG 177
Db 136273 GGTCTAGCGGGACTGTGCGAGCCAAACAGCCCGCGCGCGGAGCGCGGAGTCTG 136332
QY 178 CAGGCGCGCGGAGCGCGGAGTGGCGGAGCGCGCGCGGAGGAGGCGATCCCGCA 237
Db 136333 GAAGCGCGCGCGGAGAGGAATTCAGTGG-----CGGGCGAGCACCCAA 136380
QY 238 GCAGTCTGTGAGCGGTGTGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db 136381 GACGATCCGAGGTGACCGATGGCAGCAGAAACGCGAGGCTCCCGCGTGGCGCG 136440
QY 298 GCGCGCGCGGAGCGCGGAGGTGCGCGAGCAGCCATATACGCGCGCGCCCAAGCCC 357
Db 136441 GCGAGCGTGGCGCGCGGTGAGGGCGCGCGAGCAAGCGTACACCGCGCGCCCAAGCCC 136500
QY 358 CCTACTCGTACATCGCGCTCATCGCCATCGCCATCGCGAGTCCGCGCGCGCGCGT 417
Db 136501 CCTACTCTTACATCGCACTCATCGCCATCGCCATCGCGAGTCCGCGCGCGCGCTG 136560
QY 418 AGCTTGGCGGAGTCAACAGTACTCATGTGGGAGTTCCTTTTCCGCGCGAGCTAC 477
Db 136561 AGCTTGGCGGAGTCAACAGTACTCATGTGGGAGTTCCTTTTCCGCGCGAGCTAC 136620

Qy 478 ACAGGCTGGCGCAATCGTGGCGGACCAACCTTTCGCTCAACGACTGCTGCTCAAGGTG 537
Db 136621 ACAGGCTGGCGCAATCGTGGCGGACCAACCTTTCGCTCAACGACTGCTGCTCAAGGTG 136680
Qy 538 CTGGCGGACCCCTCGCGCCCTGGGCGAGGACAACTACTGATGCTCAACCCCAAGCAGC 597
Db 136681 CTGGCGGACCCCTCGCGCCCTGGGCGAGGACAACTACTGATGCTCAACCCCAAGCAGC 136740
Qy 598 GAGTACACCTTTCGCGGAGCGGGTCTTTCGCGCGCGCGCGCGAGCGCTCAGCAGCCGCGG 657
Db 136741 GAATACACCTTTCGCGGAGCGGGTCTTTCGCGCGCGCGCGAGCGCTCAGCAGCCGCGG 136800
Qy 658 CCGGTCCCGCGCGCGCGCGCGCGCGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 714
Db 136801 ACAGTCTCGCATCGGCGCTAGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCG 136860
Qy 715 CG 774
Db 136861 CAGCG 136920
Qy 775 CG 834
Db 136921 GGTCTCAGCG 136980
Qy 835 CCCTTCGCGAGCG 894
Db 136981 CCGTTCGCGAGCG 137040
Qy 895 GCGCGCTCG 954
Db 137041 GCTCCTCG 137100
Qy 955 CTGCTCGCGCTGCGCGTACG 1014
Db 137101 CTGCTCGCGCTGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 137160
Qy 1015 GAGGTGCCACCG 1074
Db 137161 GAGGTG---CAGCG 137217
Qy 1075 AAGCACTCCGAGGCGCGG---CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1131
Db 137218 AAGCACTTCGAGGTCCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 137277
Qy 1132 CCGGAGCGCGCTCAGCGCGCGCTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1191
Db 137278 CCGAGCGCGCTCAGCG 137337
Qy 1192 GAGCGCTCCTAGCTTGA 1209
Db 137338 GAGAGCTCTAGCTTGA 137355

RESULT 11
AC067929/c
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-403120 map 6, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
ACCESSION AC067929
VERSION AC067929.2 GI:8247824
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 182152)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-403120
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182152)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,
Meldrum,J., Mihova,T., Miranda,C., Mieng,C., Mieng,C., Morrow,J.,
Murphy,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182152)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,
Meldrum,J., Mihova,T., Miranda,C., Mieng,C., Mieng,C., Morrow,J.,
Murphy,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7655991.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center / MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10103
Center clone name: 403120
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173395 bases at least Q40
Consensus quality: 177417 bases at least Q30
Consensus quality: 179105 bases at least Q20
Insert size: 179000; agarose-ff
Quality coverage: 4.8 in Q20 bases; agarose-ff
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces

Db 9421 CCTCTCCCGCGCGCCGCCAAGCCACTCCGAGGCCCGCGCGCGCGCGCAC 9362

QY 1111 CTGTACTCCCTCGCGGTCGCCGAGCCCTGCGAGGCGGCTTAGTCCGNCGTCTCTGGC 1170

Db 9361 CTGTACTCCCTCGCGGTCGCCGAGCCCTGCGAGGCGGCTCACTCCGCGCCCTGGC 9302

QY 1171 CGCGACCTCTGTAACCGGTGAGAGCGCTCTAGCTTGA 1209

Db 9301 CGCGACCTCTGTAACCGGTGAGAGCGCTCTAGCTTGA 9263

RESULT 12

RATHPHI 1760 bp mRNA linear ROD 22-AUG-1995

LOCUS Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete

DEFINITION cds.

ACCESSION L13201 GI:951067

VERSION L13201.1

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1760)

AUTHORS Clevidence D.E., Overdier, D.G., Tao, W., Qian, X., Pani, L., Lai, E. and Costa, R.H.

TITLE Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)

MEDLINE 93248207

PUBMED 7683413

COMMENT On Aug 22, 1995 this sequence version replaced gi:550512.

Original source text: rat.

Location/Qualifiers

1..1760

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/sex="male"

/tissue_type="lung"

/dev_stage="9 week old"

/tissue_lib="Clontech"

1..1760

/gene="HNF-1"

250..1569

/gene="HNF-1"

/notes="bp 84...889 winged-helix DNA binding domain"

/codon_start=1

/product="HNF-3/forkhead homolog-1"

/protein_id="AAA74561.1"

/db_xref="GI:550513"

/translation="MGSDLGAGSDVPSPLSAAGDLSLGSDDCAANSPPAAGSAGD
LEGGERNSSGASQDDPEVTDGSRTOASVPGCAGSVGGGAGSKPVTTRPKPP
YSTIALIAIRDSAGRLTAIEINYLKGFPPFGSVTGRNVSVRNLINLDCFKV
VLDPSPKGNKWNLNSEYTFADGVFRRRRLSHRTVTSAGSGGSGPPGAG
TPQAPTAGSPTIARSPARQEGSSPAKSFSDISLSPFRSRDGLRWGCSY
PGALLPARCAPIPRFPFRRPAPVPCRSVLTARRAHAGVARGAARGAPVRGAPLH
RGPQAISSGRRRRGAPVLPPTAAHGPAAGRLRSAPSAPVLPGGDAASLTGAWSSAG
GSRVELGAPFVMEGTCHGWNAGSRAVHPFSPRNRDFFSTLHQM"

ORIGIN

Query Match 59.4%; Score 717.8; DB 10; Length 1760;

Best Local Similarity 78.7%; Pred. No. 3.1e-68;

Matches 951; Conservative 0; Mismatches 233; Indels 24; Gaps 7;

QY 8 TGGAGGTGTTCTGTCGCGCGCGCCCGAGGACAGCAGGACGTGACCTGGAGGGCG 67

Db 213 TTGGAGGTGTTCTGTCGCGCGCGCCCGAGGACAGCAGGACGTGACCTGGAGGGCG 271

QY 68 CGGCGCGCAGCAGCGCGCGCTCCCGCGTGTGCGCGCGCGGAGACGACTCCCTGGGCTCAG 127

Db 272 CGGCGAGCAGCAGCGCTGCCATCTCCACTGTCCGCGGCTGGTACGACTCCTTAGGCTCAG 331

QY 128 ATGGGGACTCGCGGCGCAA---CCGCTCCGCGCGCGCGCGCGCGAGATACGAGAGGCG 184

Db 332 ACGGGGACTGTGCAGCCAAACAGCCGCGCGCGCGCGCGCGCGGATCTGGAAGGTG 391

QY 185 ACGGGGAACAGAGTGGGAGAGGCGCGCGCGCGGAGGAGCGATCCCGGACGAGCTG 244

Db 392 CGCGCGCGAGAGGAATTCAGTGG-----CGGGGCGAGCACCCAGAGCATC 439

QY 245 CTGCAGCGTGTGTGGCGAGGCGCGAGAGCCCGGGGCGCGGGGCCAGGCGCGGGCGGG 304

Db 440 CCGAGGTGACCGATGGCAGCAGAAACGAGAGCTCCCGGCTGGGGGCGCTGTGCGCGGCGAG 499

QY 305 CGGGGACGCGGAGGCTGCACGAGCAGCAAGCATATAGCGCGCGCGCAAGCCCTTACT 364

Db 500 TGGCGCGCGGTGAGGCGCGCGCAGCAAGCCGTACACGCGCGCGGCCCAAGCCCTTACT 559

QY 365 CGTACATCGCGCTCATCGCCATCGCCATCGCGACTCGCGGCGGGCGGTGTGACGTGG 424

Db 560 CTTACATCGCACTCATCGCCATCGCCATCGCGACTCGCGGCGGACGCTGACGTGG 619

QY 425 CGGAGATCAACGAGTACTCATGGGCAAGTTCCTCTTCGCGCGGCGAGTACACGCGGT 484

Db 620 CCGAGATCAACGAGTACTCATGGGCAAGTTCCTCTTCGCGGCGGAGTACACGCGGT 679

QY 485 GGCGCAACTCGTGGCGCACAACTTTTCGCTCAACGACTGCTTCGTCGTCGTCGCGG 544

Db 680 GGCGCAACTCGTGGCGCACAACTTCGCTCAACGACTGCTTCGTCGTCGTCGCGG 739

QY 545 ACCCTCGCGGCTCGGGGCAAGCAACTACTGTGATGCTCAACCCCAACAGCGAGTACA 604

Db 740 ACCCTCGCGGCTCGGGGCAAGCAACTACTGTGATGCTCAACCCCAACAGCGAATACA 799

QY 605 CTTTCGCGAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTCAGCACCGCGCGCGGTGCC 664

Db 800 CTTTCGCGAGCGGGTCTTCGCGCGCGCGCGCGCAAGCGCTCAGCACCGCGCGGTGCC 859

QY 665 CCGCGCGCGGCTCGCGCGCGCGAGAGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCG 724

Db 860 CCGCATCGGCTACCGCGGAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 919

QY 725 CCGCGCGCGGCTCG 784

Db 920 CCAACCGCGGCTCTCCCGCAATCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 979

QY 785 CCGCGGCGCAAGTCTCCAGCTCTTCGCGCATCGACAGCATCTCTCGCAAGCGCTTCGCG 844

Db 980 CCGCGAGCAAGTCTCTCCAGCTCTTCGCGCATCGACAGCATCTCTCAGCAAGCGCTTCGCG 1039

QY 845 GCGTGGCTCAGGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 904

Db 1040 GCGCGCGGAGC-GGCACCGCGCTCTGGGGGTGACAGTACCGTGGAGCGCTGCTCCCTGCC 1098

QY 905 CGCGCTGCGCGCTTCGCGCGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 964

Db 1099 GCGCGCTGCGCGCTATCCGCGCTCTCTCCCGCGTCTCCCGGCGTGCCTCTGCTGCGCG 1158

QY 965 TCTCGGCTACG 1024

Db 1159 TCTGTGCTTACGGCGC-GGCGAGCCACGCTGTGTGCGTCTCGCGCGCGCGCGCGCG 1214

QY 1025 CGACCGCGCGCGCTCTCTGCTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1084

Db 1215 AGCCGCGCGCGCGCTGTTGTTGTTGCGCGCGCTCTCCACCGCGCGCGCGCGCGCG 1274

QY 1085 GAGCGCGCGG---CG 1141

Db 1275 GAGGTCCGGAGACCG 1394

QY 1142 TCGAGGCGCGCTTAGTCCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201

Db 1335 TGCAGGCGCGCGCGCGCGCTGCGGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCG 1394

Qy 1202 TAGCTTGA 1209

Db 1395 TAGCTTGA 1402

RESULT 13

AC012350/c

LOCUS

DEFINITION Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered

pieces.

AC012350

VERSION AC012350.3 GI:7381803

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 178168)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-16N9

Unpublished

2 (bases 1 to 178168)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Bana.N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,

Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,

Cooke,P., Dargellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

McEwan,P., McGurk,A., McKernan,K., McDonald,P., Meldrim,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tefaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X.,

Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (25-OCT-1999); Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 1, 2000 this sequence version replaced gi:6479001.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3593

Center clone name: 16_N_9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172070 bases at least Q40

Consensus quality: 174479 bases at least Q30

Consensus quality: 175597 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 176668; sum-of-contigs

Quality coverage: 6.4 in Q20 bases; agarose-fp

Quality coverage: 6.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1694: contig of 1694 bp in length

* -----

```

* 1695 1794: gap of 100 bp
* 1795 4437: contig of 2643 bp in length
* 4438 4537: gap of 100 bp
* 4538 6644: contig of 2107 bp in length
* 6645 6744: gap of 100 bp
* 6745 10092: contig of 3348 bp in length
* 10093 10192: gap of 100 bp
* 10193 11801: contig of 1609 bp in length
* 11802 11901: gap of 100 bp
* 11902 15884: contig of 3683 bp in length
* 15885 15884: gap of 100 bp
* 15885 19675: contig of 3991 bp in length
* 19676 19775: gap of 100 bp
* 19776 27913: contig of 8138 bp in length
* 27914 28013: gap of 100 bp
* 28014 34156: contig of 6143 bp in length
* 34157 34256: gap of 100 bp
* 34257 43225: contig of 9069 bp in length
* 43226 43425: gap of 100 bp
* 43426 55265: contig of 11840 bp in length
* 55266 55365: gap of 100 bp
* 55366 68306: contig of 12941 bp in length
* 68307 68406: gap of 100 bp
* 68407 82910: contig of 14504 bp in length
* 82911 83010: gap of 100 bp
* 83011 102181: contig of 19171 bp in length
* 102182 102281: gap of 100 bp
* 102282 128058: contig of 25777 bp in length
* 128059 128158: gap of 100 bp
* 128159 178168: contig of 50010 bp in length.
*
* Location/Qualifiers
*   1..178168
*     /organism="Homo sapiens"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9606"
*     /clone="RP11-16N9"
*     /clone_lib="RPC1-11 Human Male BAC"
*
* misc_feature
*   1..1694
*     /note="assembly_fragment"
*
* misc_feature
*   1795..4437
*     /note="assembly_fragment"
*
* misc_feature
*   4538..6644
*     /note="assembly_fragment"
*
* misc_feature
*   6745..10092
*     /note="assembly_fragment"
*
* misc_feature
*   10193..11801
*     /note="assembly_fragment"
*     clone_end:SP6
*     vector_side:right
*
* misc_feature
*   11902..15584
*     /note="assembly_fragment"
*
* misc_feature
*   15685..19675
*     /note="assembly_fragment"
*
* misc_feature
*   19776..27913
*     /note="assembly_fragment"
*
* misc_feature
*   28014..34156
*     /note="assembly_fragment"
*
* misc_feature
*   34257..43225
*     /note="assembly_fragment"
*
* misc_feature
*   43426..55265
*     /note="assembly_fragment"
*
* misc_feature
*   55366..68306
*     /note="assembly_fragment"
*
* misc_feature
*   68407..82910
*     /note="assembly_fragment"
*
* misc_feature
*   83011..102181
*     /note="assembly_fragment"
*     clone_end:17
*     vector_side:right
*
* misc_feature
*   102282..128058
*     /note="assembly_fragment"
*
* misc_feature
*   128159..178168
*     /note="assembly_fragment"

```


Unpublished

Baldwin, J., Barna, N., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Brown, A., Castle, A., Collangelo, R., Collins, S., Collymore, A.,
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Feirreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, F., McGuck, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:7144942.

On Jul 13, 2000 this sequence version replaced gi:714942.
All repeats were identified using RepeatMasker:
Smit, A F A. & Green. P. (1996-1997)

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L3560

Center Grove name: _____

* NOTE: This record contains 73 individual

★

RESULT 14	
AC016269/c	
LOCUS	74969 bp DNA linear HTG 13-JUL-2000
DEFINITION	Homo sapiens clone RP11-16J9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC016269
VERSION	AC016269.3 GI:9123861
KEYWORDS	HTG; HTGS PHASE0.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 74969)
AUTHORS	Barren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens, clone RP11-16J9


```

Db      38291  CCTACTCGTACATCGGCTCATCGCCATCGCATCCGAGCTCGGCGGGCGGCTTG 38232
QY      418  ACCTGCGGAGATCAACAGTACTCTCATGCGGAGTCCCTCTTTTTCGCGGCGAGCTAC 477
Db      38231  ACCTGCGGAGATCAACAGTACTCTCATGCGGAGTCCCTCTTTTTCGCGGCGAGCTAC 38172
QY      478  ACGGGCTCGGCGAATCCGTCGCGCCACACCTTTTCCTCAACGACTGCTTCGTCGAAGTG 537
Db      38171  ACGGGCTCGGCGAATCCGTCGCGCCACACCTTTTCCTCAACGACTGCTTCGTCGAAGTG 38112
QY      538  CTGCGGAGACCTCTCGGCGGCTCGGCGCAGGACAACTACTGATGCTCAACCCCAACGCG 597
Db      38111  CTGCGGAGACCTCTCGGCGG-CCTGCGGAGGACAACTACTGATGCTCAACCCCAACGCG 38053
QY      598  GAGTACACCTTCGCCGACG 616
Db      38052  GGGGGGGCTTAGAGTCG 38034

RESULT 15
AC067929
LOCUS   182152 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 6 clone RP11-403120 map 6, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
AC067929
VERSION AC067929.2 GI:8247824
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 182152)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 6, clone RP11-403120
Unpublished
2 (bases 1 to 182152)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Coddigore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, F., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182152)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Coddigore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

```

```

Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, F., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7655991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10103
Center clone name: 403 I-20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173395 bases at least Q40
Consensus quality: 177417 bases at least Q30
Consensus quality: 179105 bases at least Q20
Insert size: 179000; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 2222: contig of 2222 bp in length
* 2233: gap of 100 bp
* 2323: contig of 1463 bp in length
* 3785: contig of 100 bp
* 3883: gap of 100 bp
* 6753: contig of 2868 bp in length
* 6754: gap of 100 bp
* 6853: contig of 3183 bp in length
* 10036: contig of 100 bp
* 10137: gap of 100 bp
* 14537: contig of 4401 bp in length
* 14538: gap of 100 bp
* 14638: contig of 6540 bp in length
* 21178: gap of 100 bp
* 21278: contig of 8963 bp in length
* 30241: gap of 100 bp
* 30341: contig of 9783 bp in length
* 40123: gap of 100 bp
* 40224: contig of 7776 bp in length
* 47998: contig of 100 bp
* 48000: gap of 100 bp
* 58235: contig of 10136 bp in length
* 58335: gap of 100 bp
* 58336: contig of 8706 bp in length
* 67042: gap of 100 bp
* 67142: contig of 8780 bp in length
* 75922: gap of 100 bp
* 75922: contig of 10506 bp in length
* 76022: gap of 100 bp
* 86528: contig of 14707 bp in length
* 101334: gap of 100 bp
* 101335: contig of 16196 bp in length

```

TITLE
JOURNAL
COMMENT

Search completed: March 31, 2004, 00:21:43
Job time : 4698 secs

ORIGIN

	Query Match	20.7%	Score 250.8;	DB 2;	Length 182152;
	Best Local Similarity	71.6%;	Pred. No. 6.1e-19;		
	Matches 265;	Conservative 0;	Mismatches 102;	Indels 3;	Gaps 1;
Qy	1	ATGAAGTTGGAGTGTTCTCCTCGCGCGGCCACGGGCAACAGCAGGGCAGTGACCTG	60		
Db	29973	ATGAAGTTGGAGTGTTCTCCTCGCGCGGCCACGGGCAACAGCAGGGCAGTGACCTG	30032		
Qy	61	GAGGGCGCGGGCGGACAGCGCGCGTCCCGCTGTCCGGCGGGGAGACGACTCCCTG	120		
Db	30033	GAGGGCGCGGGCGGACAGCGCGCGTCCCGCTGTCCGGCGGGGAGACGACTCCCTG	30092		
Qy	121	GGCTCAGATGGGACTGTCGGCGCCAA---GCCGTCCGGCGGGCGCGCCAGAGATACG	177		
Db	30093	GGCTCAGATGGGACTGTCGGCGCCAAACGCCGGCGCGGGCGGGCGCGCCAGAGATACG	30152		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 21:49:37 ; Search time 502 Seconds

(without alignments)
10231.218 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209
Sequence: 1 atgaagtggagggttcgt.....tggagacgctcctagcttga 1209

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002s.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1208	99.9	1209	6	ABQ81557 Gene up-r
2	717.8	59.4	1760	7	ABT42112 Toxicity
3	442	36.6	658	6	ABQ39023 Oligonucle
C	442	36.6	658	6	ABQ39022 Oligonucle
C	305.4	25.3	658	6	ABQ39025 Oligonucle
6	305.4	25.3	658	6	ABQ39024 Oligonucle
7	201.4	16.7	2487	9	ABT76311 Prostate
8	201.4	16.7	5181	6	ABL68935 Kidney ca
9	195	16.1	1380	9	ADD29558 Human tum
10	194.6	16.1	3482	7	ABX76288 Lung canc
11	193.4	16.0	2271	6	ABZ35021 Human gen
12	187.6	15.5	1860	6	AAQ50632 Brain fac
13	184.4	15.3	1965	9	ADB31305 Testoster
14	184.4	15.3	2187	7	ABZ34783 Coding se
15	175.6	14.5	3946	2	AZ33671 Human FK
16	175.6	14.5	3946	2	AZ33671 Human FK
17	175.2	14.5	2512	5	AAZ67345 DNA encod
18	171.6	14.2	1662	2	AAZ31672 Human FK
19	171.6	14.2	1662	2	AZ338080 Human for
20	165.6	13.7	1976	2	AZ28095 Human FRE
21	161.4	13.3	2745	7	AZ220402 Oncofeta
22	161.4	13.3	2753	6	ABT03399 Ovary cel
23	160.4	13.3	1355	7	ABZ34708 Coding se

24	153.8	12.7	2934	6	ABI99306 Mouse isc
25	153.8	12.7	3289	4	AAH26570 Human win
26	153.8	12.7	6458	6	ABK48984 Genomic D
27	152.2	12.6	2106	2	AAK28103 Freac11 g
28	151	12.5	3786	5	AAK72683 DNA encod
29	150.4	12.4	2448	9	ADD29815 Human tum
30	149.6	12.4	2832	2	AAQ50631 Brain fac
31	146.2	12.1	1287	5	AAK72684 DNA encod
32	143.4	11.9	2559	9	ADD15216 Human ser
33	140.4	11.6	1422	6	ABK15142 Human hep
34	140.4	11.6	2872	6	ABL69589 Prostate
35	140.4	11.6	2872	6	ABT11015 Human bre
36	140.4	11.6	2872	7	ACC50184 Breast ca
37	140.4	11.6	3098	5	ABV25148 Human pro
38	140.4	11.6	11613	6	ABK15141 Human hep
C	139	11.5	1711	6	ABQ50740 Oligonucle
40	139	11.5	1711	6	ABQ50741 Oligonucle
41	136.2	11.3	812	6	ABQ61118 Transcrip
42	136	11.2	441	7	ABX41089 Bovine ES
43	134.8	11.1	4450	4	AAK90986 Human dig
44	134.8	11.1	4450	5	AAK32021 Human liv
45	134.8	11.1	4450	6	ABN90376 Human liv

ALIGNMENTS

RESULT 1

ABQ81557
ID ABQ81557 standard; cDNA; 1209 BP.

XX
AC ABQ81557;

XX
DT 30-DEC-2002 (first entry)

XX
DE Gene up-regulated in metastatic colorectal cancer.

XX
KW Colorectal cancer; metastasis; differential expression; cytostatic;

XX
KW diagnosis; gene therapy; vaccine; gene; ss.

XX
OS Homo sapiens.

XX
PH Key Location/Qualifiers

FT CDS 1..1209

FT /*tag= a

FT /product= "Metastatic colon cancer polypeptide"

XX
PN WO200268677-A2.

XX
PD 06-SEP-2002.

XX
PF 27-FEB-2002; 2002WO-US006001.

XX
PR 27-FEB-2001; 2001US-0272206P.

XX
PR 02-APR-2001; 2001US-0281149P.

XX
PR 17-APR-2001; 2001US-0284555P.

XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.

XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX
XX Mack DH, Markowitz SD;

XX
XX WPI; 2002-698677/75.

XX
XX P-PSDB; ABP54691.

XX
XX New genes that are up- or down-regulated in colorectal cancer, useful for

XX
XX diagnosing colorectal cancer in a subject, or for identifying modulators

XX
XX of colorectal cancer-associated proteins and genes for treating

XX
XX colorectal cancer.

XX
XX Claim 5; Page 252; 260pp; English.

XX
XX The present sequence is the nucleotide sequence of a human gene that

XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
PT modeling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
XX
XX Example 4; Page; 446pp; English.
XX
XX The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the
CC progression of a toxic effect of a compound, predicting the renal
CC toxicity of a compound, or identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;
Query Match 59.4%; Score 717.8; DB 7; Length 1760;
Best Local Similarity 78.7%; Pred. No. 3.2e-87;
Matches 951; Conservative 0; Mismatches 233; Indels 24; Gaps 7;
8 TGGAGTGTTCCTCCGCGGCGCCACGCGGCAAGCAGCGGCGATGACCTGGAGGGCG 67
213 TTGGAGTGTTCCTCCGCGGCGCCACGCGGCAAGCAGCGGCGATGACCTGGAGGGCG 271
68 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 127
272 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 331
128 ATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 184
332 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 391
185 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 244
392 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 439
245 CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 304
440 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 499
305 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 364
500 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 559
365 CGTACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 424
560 CTTACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 619
425 CGGAGATCAACGAGTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 484
620 CCGAGATCAACGAGTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 679
485 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 544
680 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 739

QY 545 ACCCTCGCGGCGCTTGGGCGAAGACAACTACTGGATGCTCAACCCCAACAGCGAGTACA 604
DB 740 ACCCTCGCGGCGCTTGGGCGAAGACAACTACTGGATGCTCAACCCCAACAGCGAGTACA 799
QY 605 CTTTCGCGGCGGCGCTTTCGCGCGCGCGCGCGCGCGCTTCAGCCACCGCGCGCGCTCC 664
DB 800 CTTTCGCGGCGGCGCTTTCGCGCGCGCGCGCGCGCGCTTCAGCCACCGCGCGCGCTCC 859
QY 665 CCGCGCGCGGCGCTTCGCGCGCGCGCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCG 724
DB 860 CCGCATCGGCGCTACGCGCGGAGGAGCCACCGCGGCGCGCGCGCGCGCGCGCGCGCG 919
QY 725 CCGCGCGCGCGCGCTTCG 784
DB 920 CCAACCGCGCGCTTCCTCCCAATCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCG 979
QY 785 CCGCGGCGGAGTTCCTCAGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 844
DB 980 CCGCGGAGCAAGTTCCTCAGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1039
QY 845 GCCGTGCGCTCAGGCGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 904
DB 1040 GCCCGCGGCG 1098
QY 905 CGCGCGTTCGCGCGCTTCG 964
DB 1099 CGCGCGTTCGCGCGCTTCG 1158
QY 965 TCTCGCGGTACG 1024
DB 1159 TCTGTGTTACG 1214
QY 1025 CGACCG 1084
DB 1215 AGCG 1274
QY 1085 GAGCG 1141
DB 1275 GAGTTCGCGGAGACCG 1334
QY 1142 TGACGCGCGCGCTTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1201
DB 1335 TGCAGCG 1394
QY 1202 TAGCTTGA 1209
DB 1395 TAGCTTGA 1402
RESULT 3
ABQ39023
ID ABQ39023 standard; DNA; 658 BP.
XX AC ABQ39023;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25614.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP010074.
XX PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.
XX (EPiG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 658 BP; 173 A; 310 C; 108 G; 67 T; 0 U; 0 Other;
Query Match 36.6%; Score 442; DB 6; Length 658;
Best Local Similarity 82.7%; Pred. No. 1.6e-50;
Matches 505; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 599 AGTACACCTTCGGCGAGCGGTCTTCGGCGCGCGCGCGAGCGCTCAGCGACCGCGCG 658
DB 2 AATACACTTCGGCGAGCGAATCTTCGGCGCGCGCGCGAAGCGCTCAGCGACCGCGCG 61
QY CGGTCCCG 718
DB 62 CGATCCCG 121
QY 719 CCG 778
DB 122 CCG 181
QY 779 CCAGCG 838
DB 182 CCAACCG 241
QY 839 TCCG 898
DB 242 TCCG 301
QY 899 CCTCG 958
DB 302 CCTACCG 361
QY 959 TCCG 1018
DB 362 TACCGAA 421
QY 1019 TGCCACCGAA 1078
DB 422 TACCACCGAA 481

QY 1079 CACTCCGAGCG 1138
DB 482 CACTCCGAGCG 541
QY 1139 CCCTGCGAGCG 1198
DB 542 CCCTGCGAGCG 601
QY 1199 TCCTAGCTTGA 1209
DB 602 TCCTAACCTAA 612
RESULT 4
ID ABQ39022/c
XX ABQ39022 standard; DNA; 658 BP.
AC ABQ39022;
DT 12-JUL-2002 (first entry)
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25613.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
OS Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP010074.
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX (EPiG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 658 BP; 67 A; 108 C; 310 G; 173 T; 0 U; 0 Other;

```
Query Match      36.6%; Score 442; DB 6; Length 658;
Best Local Similarity 82.7%; Pred. No. 1.6e-50;
Matches 505; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 599 AGTACACCTTCGCGACGGGGTTCCTCCGCGCGCGCGCAAGGCGCTCAACACCGCGCGC 658
DB 657 AATACACCTTCGCGACGAAATCTTCGCGCGCGCGCAAGGCGCTCAACACCGCGCGC 598
QY 659 CGGTCCCGCGCGCGCGCTGCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGC 718
DB 597 CGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 538
QY 719 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 778
DB 537 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 478
QY 779 CGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 838
DB 477 CCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 418
QY 839 TCGCAGCGCGCTCGCTTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 898
DB 417 TCGCGAACCGCGCGCTCAAAACACGACCGCGCGCGCGCGCGCGCGCGCGCGCGC 358
QY 899 CTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 958
DB 357 CCTACCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 298
QY 959 TCGCGCTGTGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1018
DB 297 TACCGCTGTACGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 238
QY 1019 TCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1078
DB 237 TACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 178
QY 1079 CACTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1138
DB 177 CACTCGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 118
QY 1139 CCGTGCAGCGCGCTTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1198
DB 117 CCTACAAACGACCTCAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 58
QY 1199 TCGTAGCTTGA 1209
DB 57 TCGTAGCTTAA 47

RESULT 5
ABQ39025/c
ID ABQ39025 standard; DNA; 658 BP.
XX
AC ABQ39025;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25616.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-BP010074.
XX
```

```
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 658 BP; 269 A; 222 C; 108 G; 59 T; 0 U; 0 Other;

Query Match      25.3%; Score 305.4; DB 6; Length 658;
Best Local Similarity 68.6%; Pred. No. 2.1e-32;
Matches 420; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 598 GAGTACACCTTCGCGACGGGGTTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGC 657
DB 658 GAGTATATTTTCGTCGACGGGGTTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGC 599
QY 658 CGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 717
DB 598 TCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 539
QY 718 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 777
DB 538 TTGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 479
QY 778 GCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 837
DB 478 GTTAGTTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 419
QY 838 TTGCGAGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 897
DB 418 TTTCGTAGTCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 359
QY 898 CCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 957
DB 358 TTTCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 299
QY 958 CTGCGCGCTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 1017
DB 298 TTTCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 239
QY 1018 GTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1077
DB 238 GTGTATCATCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGC 179
```


PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 25-SEP-2000; 2000US-023537P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 28-SEP-2000; 2000US-0236422P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237385P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 PI WPI; 2002-198264/24.
 DR
 DR
 DR
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 PS
 PS Claim 1; SEQ ID NO 7272; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in AB461664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour
 XX
 SQ Sequence 5181 BP; 1167 A; 1450 C; 1340 G; 1224 T; 0 U; 0 Other;
 Query Match 16.7%; Score 201.4; DB 6; Length 5181;
 Best Local Similarity 52.1%; Pred. No. 9.1e-19;
 Matches 574; Conservative 0; Mismatches 512; Indels 16; Gaps 5;
 83 CGCCGTCCTCCGCGCGGCGGAGACGACTCCCTGGGCTCAGATGGGAGCTGCGCG 142

Db 2790 CGACGACGAGGCGCGGTGCGGGCCCGCGCTGGCTGTCCCGCGAGCGGCGCGCG 2849
 Qy 143 CC--AAGCGTCGCGCGCGCGCGCGACAGAGATACGACAGGCGACGCGCAACAGAGTGC 200
 Db 2850 CGCGCGCTGTACCGCGGAGGAGAGAGCTGAGGATCTGGAGGAGAGAGACGACGA 2909
 Qy 201 GGGAGGCGGCGCGCGCGGAGGAGCGATCCCGGACAGAGCTGCTCAGCGTGTGGC 260
 Db 2910 TGACATCTGTGGCGCGCGCTGCTGGGGCTCCCGCGCGCGCGCGCGCGCGCG 2959
 Qy 261 GGAGGCGCGAGCGAGG 320
 Db 2970 GCGGCGGCGAGAGCGCGGTGGGGCGG--CGGCGCGCGCGCGCGCGCGCGCGGAG 3026
 Qy 321 TGCACGACGAAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA 380
 Db 3027 CGCGGTAGCGTCA 3086
 Qy 381 CGCATGCGCATCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTA 440
 Db 3087 CACTATGCGCATCTGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGT 3146
 Qy 441 CCTCATGGGCAAGTTCCCTTTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
 Db 3147 CATCAGCGCGCTTCCCTACTACCGGAGAGTTCCTCCCGCTGGCAGAACAGCATCG 3206
 Qy 501 CCACAACTTTCGCTCAACGACTTCTGTCAGAGTCTGCGGACCGCTCGGGCGCGCTG 560
 Db 3207 CCACAACTCTCGCTCAACGACTTCTGTCAGAGATCCCGCGCGAGCGCC--GGCAACCC 3263
 Qy 561 GGGCAAGGACAACTACTTGGATGCTCAACCGAGGTACACCTTCGCGCGAGCGGGT 620
 Db 3264 GGGCAAGGCAACTACTTGGAGCTGACCGGAGTCCCGCGACATGTTGACAAACGCGAG 3323
 Qy 621 CTTCCG 680
 Db 3324 CTTCTCGCGCGAGGAGCGCTTCAAGCGCGAGCGCGCTGCTCCACCGCGCGCGCG 3383
 Qy 681 GCCGAGGAGCTC 740
 Db 3384 CGCGAGTCTGTGTGTGCG 3443
 Qy 741 GCCCGCGATGCGTGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
 Db 3444 CGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 3497
 Qy 801 CAGCTCCTTCCCATCGACAGCATCTTGGCAAGCGCGCTTCCGCGAGCGCTCGCGCTCAG 860
 Db 3498 CGGCGCGCTACGGCTGCGGCTACGCGCTTACGCGCGCGCGCTTACGCGCGCGCTT 3557
 Qy 861 CACGTT 920
 Db 3558 CTTCCG 3617
 Qy 921 CCGCGCGCTCTCCCG 980
 Db 3618 GCCCG 3677
 Qy 981 GGGCGAGCG 1040
 Db 3678 GCACCGCT--CGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3735
 Qy 1041 CTTGCTGACCTTCCCG 1100
 Db 3736 GCGCGCGCGCGCTCAGCGCTGCGCGCGCTGCGCGCTTCTCCATCGAGAGCATCAT 3795
 Qy 1101 CGGCGCGCGCGCTTACTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1160
 Db 3796 AGCTTGGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3855
 Qy 1161 NCGTCTGGCGCGCGCGCTGCG 1182

Db 3856 TCGCCCTCGCCCTCGCCGGTGG 3877

RESULT 9

ADD29558

ID ADD29558 standard; mRNA; 1380 BP.

XX AC

XX ADD29558;

XX DT

XX 15-JAN-2004 (first entry)

XX DE Human tumour suppressor mRNA SEQ ID NO:7.

XX KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.

XX OS Homo sapiens.

XX PN WO2003056201-A2.

XX PD 17-JUL-2003.

XX PF 31-DEC-2002; 2002WO-US041825.

XX PR 31-DEC-2001; 2001US-0345317P.

XX PA (QUAR-) QUARK BIOTECH INC.

XX PA (CLEV-) CLEVELAND CLINIC FOUND.

XX PI Feinstein E, Gudkov AV;

XX PI WPI; 2003-598393/56.

XX PT Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.

XX PS Disclosure; SEQ ID NO 7; 272pp; English.

XX CC The invention relates to a novel method for diagnosing a cancer in a subject. The method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of cancer is indicative of cancer. The polypeptide is selected from any of the polypeptides encoded by the polynucleotides listed in the specification and polypeptides which are at least 70% homologous to the polypeptides. The method of the invention has cytostatic activity, and may have a use in gene therapy. The method is useful in identifying markers specific for one or several types of cancer, depending on the tissue origin, which may be used in numerous diagnostic and prognostic applications as well as cancer type-specific targets for therapeutic intervention. The compounds that modulate the activity of a tumour suppressor gene are useful in the treatment of cancer or as anti-cancer drugs. The present sequence represents a polynucleotide of the invention.

XX SQ Sequence 1380 BP; 216 A; 496 C; 500 G; 168 T; 0 U; 0 Other;

Query Match 16.1%; Score 195; DB 9; Length 1380;

Best Local Similarity 53.1%; Pred. No. 8.3e-18;

Matches 533; Conservative 0; Mismatches 455; Indels 15; Gaps 5;

QY 2 TGAAGTTGAGGTGTCGTCCTCGCGGGCCAGCGGGACAGGAGGCGGTGACCTGG 61

Db 74 TGGACATCATGTGTGGCGAGGCGACGCGGGCTGGAAGAGAGACGACGACGCGAG 133

QY 62 AGGGCGCGGGCGGACGACGCGCGCTGCCCG--CTGTGGCGCGGGAGACGACTCCC 118

Db 134 GTTGGATAGCCCCCGGGCGCGCGAGCTGCGCTGGACGAGCGGACGAGTGGCCCC 193

QY 119 TGGGCTCAGATGGGACTCGCGGGCCAGCCGCTTCGCGGGCGGGCGGCCAGATAGCG 178

Db 194 CGCGCGGACCCCATCAGCGACAGCTCAGCGCCGCCCGCCACGAGCGCCCTGATGGCCCA 253

QY 179 AGGCGGACGGCGAACAGAGTGCGGGAGCGCGCGCGCGGAGGAGGCGATCCCGGCG 238

Db 254 AGGAGGCGGCGGAGCCGGGCGCGAGCTGGGCGCGCGCGCGCGCGGAGCGGACG 313

QY 239 CAGCTGCTGACGGTGGTGGCGGAGGCGCGGAGGCGCGGGCGCGGGCGCGGCGG 298

Db 314 GCTCAAGGGCGGTGTTGGCGGAGGAGGCGCGGAGGCGCGGGCGCGGCGCGCGG 373

QY 299 GCGGCGGGGAGCGCGAGGCGTGCAGCGAGCAAGCCATATACGCGCGCGCCCAAGCCCC 358

Db 374 GCAGCGTTCGGCGGAGGCGCTGCCCCGAGCAAGCCCAAGAACAGCCTAGTGAAGCGCG 433

QY 359 CTTACTGTCATCGCGCTCATCGCCATGGCCATCGCGACTCGGGGGCGGGCGGCTTGA 418

Db 434 CTTACTGTCATCGCGCTCATCGCCATGGCCATCGTCAGAGCCCGCGAGCAAGAGCTGA 493

QY 419 CGTTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTTCCCTTTTCCGCGGCGAGTACA 478

Db 494 CCCTGAGCGGCATCTGGGAGTTTCATCAGCAACCGCTTCCCTTACTACAGGAGAGTTCC 553

QY 479 CGGCTGGCGCAACTCGTGGCGCACAACTTTGGTCAACGACTGCTTGTCAAGGTGC 538

Db 554 CGGCTGGCGCAACAGCATCCGCCACAACTCTCACTCAACGACTGCTTGTCAAGATCC 613

QY 539 TGGCGGACCCCTCGCGGCGCTGGGGCAAGGACAATCTACTGGATGCTCAACCCCAAGCG 598

Db 614 CCGCGGAGC--CGGGCAACCCGGGCAAGGCACTACTGGACCTCGGACCCGCGAGTCCG 670

QY 599 AGTACACCTTTCGCGGAGCGGGTTCCTCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCG 658

Db 671 AGGACATGTTTCGACCAACGCGAGCTTCTCGCGGCGCGGAAACGCTTCAAGCGCACCG 730

QY 659 CGGTCCCGCGCGCGGCTCGCGCGCGCGGAGGCGCGCGGCGCTCCCGCGCGCGCGCGCG 718

Db 731 AGGAGCACCTTGGCGGAGCAGCGCGGCTCATGATGAGAGCTTCGGGCGTTACAGCTGG 790

QY 719 CCGCGCGCGCGCGCGCGCTCGCGCGCGCATCGCTCGCGCGCGCGCGCGCGCGCGCGCG 778

Db 791 CGGCGCGCGCGCGCGCGCGCGGACCC--TACGCGCGCGCGCTTACGCGCTGCACCTGCG 846

QY 779 CCGCGCGCGCGCGGAGTTTCCAGCTCTTCGCGCATCGACAGC--ATCTGCGCGAGCG 836

Db 847 GCGCGCGCGCGTGCCTATTTCGACCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 906

QY 837 CTTCCGCGAGCGCTCGCTCAGGAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 896

Db 907 CTCAGTACCGGTACGCGCTCGCGCGGTGGCACCGGTGCTGCTCCCGCTGTGCGCGC-- 964

QY 897 GCCTGCGCGCGCGCTGCGCGCGTTCGCGCGCTTCCTCCCGCGCGCGCGCTTCAGGCGCCT 956

Db 965 -TGCTGCGCTTCGGCGGAGCTGGGCGCGCAAGCGCGCGCGCTTCGGGCTCAGAGCTCGGCGCG 1023

QY 957 GCTGCGCGCTCTGCGCGTACGCGCGCGCGCGGCGCGCGCGCGCGCGCGCTG 999

Db 1024 GGCTGCGAGTGCAGCTCAATAGCTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066

RESULT 10

ABX76288

ID ABX76288 standard; DNA; 3482 BP.

XX AC ABX76288;

XX DT 02-APR-2003 (first entry)

XX DE Lung cancer-associated polynucleotide #152.

XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
XX antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX interstitial pulmonary fibrosis; asthma; bronchiectasis.

OS	Unidentified.
XX	
PN	W0200286443-A2.
XX	
PD	31-OCT-2002.
XX	
PF	19-APR-2002; 2002WO-US012476.
XX	
PR	19-APR-2001; 2001US-0284770P.
PR	10-MAY-2001; 2001US-0290492P.
PR	09-NOV-2001; 2001US-0339245P.
PR	13-NOV-2001; 2001US-0350666P.
PR	29-NOV-2001; 2001US-0334370P.
PR	12-APR-2002; 2002US-0372248P.
XX	(EOSB-) EOS BIOTECHNOLOGY INC.
PA	
XX	Aziz N, Murray R;
PI	
XX	WPI; 2003-093161/08.
DR	P-PDSB; ABU56559.
XX	
PT	Detecting a lung cancer-associated transcript in a cell from a patient
PT	for treating lung cancer, by contacting a biological sample from the
PT	patient with a polynucleotide that exhibits increased or decreased
PT	expression in lung cancer.
XX	
PS	Claim 22; Page 303-304; 453pp; English.
XX	
CC	The invention relates to a method for detecting a lung cancer-associated
CC	transcript in a cell from a patient, comprising contacting a biological
CC	sample from the patient with a polynucleotide that selectively hybridises
CC	to a sequence that is at least 80 % identical to a gene that exhibits
CC	increased or decreased expression in lung cancer samples. Lung cancer-
CC	associated polynucleotides and polypeptides are used for identifying a
CC	compound that modulates a lung cancer-associated polypeptide, for
CC	inhibiting proliferation of a lung cancer-associated cell to treat lung
CC	cancer in a patient and for treating a mammal having lung cancer by
CC	administering a modulatory compound identified. The methods are useful
CC	for treating lung cancer, such as small cell lung cancer, non-small cell
CC	lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC	emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC	hyper-sensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC	bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC	for diagnostic purposes and as targets for screening for therapeutic
CC	compounds that modulate lung cancer, such as antibodies. Sequences
CC	CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
CC	invention
XX	
SQ	Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;
	Query Match 16.1%; Score 194.6; DB 7; Length 3482;
	Best Local Similarity 57.9%; Pred. No. 7.8e-18;
	Matches 385; Conservative. 0; Mismatches 274; Indels 6; Gaps 2
Qy	143 CCAAGCCGTCGCCGGCGCGCGGCACAGATACGACGGGCGACGGCGAACAGATGCGCG 202
Db	611 CCCCGACACCGCGGGGATCCAGACC CGGGGTGCGGACGCGCGCATGACTGCGCG 670
Qy	203 GAGGCGGGCGCGCGCGGAGGAGGCGATCCCGGCGAGCAGTGTGTGACAGCGGTGGTCGCG 262
Db	671 AGAGCGGGCGCGCGCGCGCGCACCG-- --GAGTGCTGGCTACGCTGAAGAGACGGCG 727
Qy	263 AGGCGCGGAGGCGGGCGCGCGCGCGCGACGCGCGGGCGCGCGGAGCGCGGAGGTG 322
Db	728 GCAGAGCGCAGCAGCGGGCGCGGGTCTCCAGGGGAGGCCACCGGGCGCGGGCGGGCG 787
Qy	323 CAGCGAGCAGCAATATACGCGCGCGCGCCAAAGCCCCCTACTCGTACATCGCGCTCATCG 382
Db	788 GGGCGCGCAGCGCCCGCTGACGCGCGGGAGACCGCCCTACAGCTACATCGCGCTCATCG 847
Qy	383 CCATGGCCCATCCGCGACTCGCGGGCGCGCGCGTTGACGTGGCGGAGATCAACAGTACC 442

CC	epithelium, bronchial epithelium, mammary epithelium, prostate
CC	epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC	small airway epithelium, renal epithelium, umbilical artery smooth
CC	muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC	dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC	aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
CC	bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
CC	osteoblasts or prostate stromal cell. The gene expression profile is used
CC	for determining the level of RNA expression for a sample, determining the
CC	phenotype of a cell and distinguishing cell types. The gene or a protein
CC	expression profile is useful in identifying disease pathologies involving
CC	alterations of gene expression. The assessment of expression profiles may
CC	provide meaningful information with respect to tumour type and stage.
CC	Treatment methods, and prognosis. The gene or protein expression profile
CC	may also be used for creating microarrays. The microarray is useful for
CC	genetic and physical mapping of genomes, DNA sequencing, genetic or
CC	medical diagnosis, genotyping of organisms, confirming cell or tissue
CC	identifications and in identifying promising antibiotics, antiviral or
CC	antifungal agents
XX	
SQ	Sequence 2271 BP; 392 A; 794 C; 705 G; 380 T; 0 U; 0 Other;
Query Match 16.0%; Score 193.4; DB 6; Length 2271;	
Best Local Similarity 52.8%; Pred. No. 1.2e-17;	
Matches 490; Conservative 0; Mismatches 427; Indels 11; Gaps 3;	
QY	255 GGTGGCGGAGCGCGGAGCCGGGGCGGGCGGCGGCGGCGGCGGCGGAGCGG 314
DB	440 GGCCTCCGCGCGGGGCGAGAGCCGTGGGGGCGGCGGCGGCGGCGGCGGCGG 499
QY	315 CGAGGGTGCGAGCAAGCCATATACGGCGGCGGCCAACAGCCCCCTACTCTGACATCGC 374
DB	500 CGGAGCGCGGTAGCGCGGCCAAGAACCCGCTGGTGAAGCGGCCCTACTCGTATATCGC 559
QY	375 GCTCATGCGCATGGCGATCCCGGACTCGGCGGGCGGGCGGCTTGACGCTGGCGGNGATCAA 434
DB	560 GCTCATCACTATGGCCATCTCTGCAGAGCCCAAGAACGGCGTACGCTGAGCGAGATCTG 619
QY	435 CGAGTACTCATGGCGAAGTTCCTCTTTTCCGCGGCGAGCTACACGGGCTGGCGCAACTC 494
DB	620 TGAGTTTATAGCGGCGCTTCCCTACTACCGGAGAGATTCCCGCTGGCAGAACAG 679
QY	495 CGTGGCGCAACCTTTTCGTTCACAGACTGTTTCGTCAAGTGTCGTGGCGACCCCTCGCG 554
DB	680 CATCCGCGCAACCTCTCGTTCACAGACTGTTTCGTCAAGATATCCCGCGGAGGCC--GG 736
QY	555 GCCCTGGGGGCAAGACAACACTACTCGATGCTCAACCCCAACAGCGAGTACACCTTCGCCGA 614
DB	737 CAACCCCGGGCAAGGGCAACTACTCGAGCGCTGGACCCGGAGTCCCGCGACATGTTTCGA 796
QY	615 CGGGGTCTTTCGCGCGCGCGCCGCAAGCGCTCAGCCACCGCGCGCGCTCCCGCGCCCGCG 674
DB	797 CGGCAGCTTCTCTGCGCGGAGGAGCGCTTCAAGCGGACGCGCTGCTCCACCCCAACGC 856
QY	675 GCTCGGCGCGAGAGAGCCCGGGCTCTCCCGCGCGCCCGCGCGCGCGCGCGCGCC 734
DB	857 CGCGGCGCGCGAGTCTCTGTGCTGCGCGCGGAGCCGACGAGGGGCGCGGGCGACCC 916
QY	735 GGCCTCGCCCGCATGGCTCGCCCGCCCGCGCAGGAGGAGCGGCGCCAGCGCCCGGGCGCAA 794
DB	917 GGCAGCGCGCGCGCGCTCTTCCCGCGCGCGCCCGCGCC-----CCGCGATGCTTA 970
QY	795 GTTCTTCAGTCTTTGCGCATTCGACAGCATCTCTGGGCAAGCCCTTCGCGAGCGGTGCGCT 854
DB	971 CGGCTACGGCCCTTACGGCTCGGCTACGGCTCGACGTGCGGCTTACGCGCGCGCCCTC 1030
QY	855 CAGGGACAGGCGCCCGGGAGCGAGCTTCACTGGGGGCGCGCGCCCTGTGCGCGCGTGGCC 914
DB	1031 GGCCCTCTTTCGCGCGCGAGCGGCGCGCGCGCGCGCGCTTCCACCGGCACTGCC 1090
QY	915 CGCGTTCCCGCGCTCTCTCCCGGGGCGCCTTCGAGGGCCCTGTGCGGCTCTGCGGCTA 974
DB	1091 CCCGCGCGCCCGGCAACGCGACGCGCGGCGCGCGCGAGCTGGCCCGGACCGGCTTCGGCTA 1150

XX OS Homo sapiens.
XX PN WO9952415-A2.
XX PD 21-OCT-1999.
XX PF 14-APR-1999; 99WO-US008159.
XX PR 15-APR-1998; 98US-0081870P.
XX PR 22-MAY-1998; 98US-00083351.
XX PA (IOWA) UNIV IOWA RES FOUND.
XX PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;
XX DR WPI; 1999-620257/53.
XX DR P-PSDB; AAY43260.
XX PT New isolated human forkhead transcription factor gene, FKHL7, used to
XX PT develop products for the diagnosis, prognosis, monitoring, prevention or
XX PT treatment of congenital heart disease.
XX PS Claim 11; Fig 1; 98pp; English.
XX CC This sequence encodes the human forkhead transcription factor gene,
XX CC designated FKHL7, of the invention. FKHL7 can be used in a novel method
XX CC for treating or preventing the development of a congenital heart disease
XX CC (CHD) in a subject. The FKHL7 sequences can be used for diagnosis,
XX CC prognosis, monitoring, prevention and treatment of CHD. They can also be
XX CC used for the production of transgenic animals and drug screening
XX SQ Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 U; 0 Other;

Query Match 14.5%; Score 175.6; DB 2; Length 3946;
Best Local Similarity 50.1%; Pred. No. 2.5e-15;
Matches 552; Conservative 0; Mismatches 534; Indels 16; Gaps 4;

QY 23 CTCGCGGCGCCACAGGCGACATGACCTGAGGGCGCGGGCGGAGCGACG 82
Db 372 CTCGCGGCGCGGACTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCG 431

QY 83 CGCGTCTCCCGCTGCGGCGCGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCG 142
Db 432 GTGGGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 490

QY 143 CCAAGCGCTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202
Db 491 CCGTGTCCAGCCCACTCCCTGSGAGTGTGCTTACCTCGGCGCGAGCAGACTACT 550

QY 203 GAGCGGCGCGGCGCGGAGAGCGATCCCGGCGAGCTGTGTCAGCGGTGTGCGCG 262
Db 551 ACCGCGGCG 610

QY 263 AGGCGGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 322
Db 611 TGTACTCGACCCCTGCGACCGCGAGAGTACCGCGCGCGCGCGCGCGCGCG 670

QY 323 CACGACGACGACATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376
Db 671 CCTACACG 730

QY 377 TCATCG 436
Db 731 TCATCAGCATGGCGCATCCAGAACCGCGCGCGCGCGCGCGCGCGCGCGCG 790

QY 437 AGTACCTCATGGGCAAGTTCCTCTTTTTCGCGCGCGCGCGCGCGCGCGCG 496
Db 791 AGTTTCATCATGACCGCTTCCCTTCTACCGGAGCAACAGCAGGCGTGGCAG 850

QY 497 TGGCGCACAGCTTTCGCTCAACGACTGCTTCGTCAAGGTGCTGCGCGCGCG 556
Db 851 TCCGCCCAACCTCTCGCTCAACGAGTCTTCGTCAAGGTGCTGCGCGCGCG 910

QY 557 CCTGGGGCAAGGACAACTACTGGATGTCAACCCCAACAGCGAGTACACCTTCGCGGACG 616
Db 911 C---GGGCAAGGCGAGCTTACTGGACGCTGGAGCCCGGACTCTCTACACATGTTTCGAGAAG 967

QY 617 GGGTCTTCCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGCTCCCGCGCGCGCGCG 676
Db 968 GCAGCTTCTGCGGCGCGCGCGGCTTCAAGAGAGAGAGCGCGGTGAGAGCAAGGAGG 1027

QY 677 TCGCGCGCGGAGAGGCC-----CGGGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 730
Db 1028 AGAAGGACAGGCTGCACCTCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087

QY 731 CCGCGCGCTCCCG 790
Db 1088 CGCGGAGCAGCG 1147

QY 791 GCAAGTTCTCCAGCTCTCTTCGCCATCGACAGCATCTCTCGCAAGCGCTTTCGCGAGCCGTC 850
Db 1148 ACATCAAGACCGAGAACGTTAGTGCCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1207

QY 851 GCCTCAGGAGACAGCG 910
Db 1208 CCTGGGCGAGCG 1267

QY 911 TGCCCGCGTTCGCCCGCGCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
Db 1268 GCAGCTGTTCAGCGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327

QY 971 CGTACGCGCGGCGAGCG 1030
Db 1328 TGGACGCTGCGGATTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1387

QY 1031 CGCGCGCGCTCTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1090
Db 1388 GCCAGGGCTTCAGCGTGGACACATCATGACGTGCTGCGGGGCTCGCGCGAGACGCGG 1447

QY 1091 CG 1112
Db 1448 CGCGGAGCTCAGCTTCGCGCGCT 1469

Search completed: March 30, 2004, 23:01:41
Job time : 508 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	271.8	22.5	320	4	US-09-833-681-1266
2	187.6	15.5	1860	2	US-08-331-644-3
3	187.6	15.5	1860	5	PRT-US93-04102-3
4	184.4	15.3	1965	4	US-09-220-132-21
5	175.6	14.5	3946	3	US-09-083-351-1
6	175.6	14.5	3946	3	US-09-083-352-1
7	171.6	14.2	1659	3	US-09-083-351-3
8	171.6	14.2	1659	3	US-09-083-352-3
9	152.8	12.6	2830	1	US-07-882-392-1
10	152.8	12.6	2830	2	US-08-331-644-1
11	152.8	12.6	2830	5	PCT-US93-04102-1
12	134.6	11.1	1309	4	US-09-976-594-927
13	130.2	10.8	1155	2	US-08-331-644-4
14	130.2	10.8	1155	5	PCT-US93-04102-4
15	117.6	9.7	5080	4	US-09-976-594-495
16	110.6	9.1	1634	3	US-09-087-134-13
17	110.6	9.1	1793	3	US-09-113-309-1
18	110.6	9.1	1793	3	US-09-521-109-1
19	110.6	9.1	1793	4	US-09-562-332-1
20	104.4	8.6	2561	4	US-09-616-289-48
21	91.2	7.5	12001	1	US-08-458-568A-11
22	88	7.3	2561	4	US-09-616-289-48
23	86	7.1	4403765	3	US-09-103-840A-2
24	85.4	7.1	152331	3	US-09-128-155-16
25	85	7.0	12425	4	US-09-616-289-50
26	84.8	6.9	4411529	3	US-09-103-840A-1
27	83.6	6.9	4403765	3	US-09-103-840A-2
					Sequence 1266, Ap
					Sequence 3, Appl
					Sequence 3, Appl
					Sequence 21, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 3, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 127, App
					Sequence 4, Appl
					Sequence 4, Appl
					Sequence 495, App
					Sequence 13, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 1, Appl
					Sequence 48, Appl
					Sequence 11, Appl
					Sequence 48, Appl
					Sequence 2, Appl
					Sequence 16, Appl
					Sequence 50, Appl
					Sequence 1, Appl
					Sequence 2, Appl

Query Match 15.5%; Score 187.6; DB 5; Length 1860;
Best Local Similarity 58.3%; Pred. No. 3.6e-22;
Matches 394; Conservative 0; Mismatches 264; Indels 18; Gaps 3;
QY 8 TGGAGGTGTTGTCCTCGCGGCGCCACCGGGAGACAGCAGGCGAGTACCTGGAGGGCG 67
Db 421 TCGACGTGTTGGGGAGGCGAGGACGACGAGGAGGAGGACGATGACGACGAGGGCG 480
QY 68 CGGGCGGACGACGCGCTCCCGCTGTCGGCGCGGAGACGACTCCCTGGGCTCAG 127
Db 481 GCGGCGGCGCGCGCGGGGGTCCCGTTCGCGAGCTCGGCCACGCGGCGAGGCGT 540
QY 128 ATGGGAGTGCAGCGGCGCAAGCCGTCGCGCGGCGGCGGCCCGCCAGATACGACGAGGCGAGC 187
Db 541 CTTACCGCGGAGGTGATC-----TCGAGGACCTGGAGGAGGAGGACGACG 588
QY 188 GCGAACAGAGTGCAGGAGCGGCGCGGCGCGGAGGAGCGATCCCGCGAGCGTGTG 247
Db 589 ATGACCTGTGTGCGCCCCCGCGCGCGCTGTCGCCCGCGCTCCCGGCTCTGCGCGCG 648
QY 248 CAGCGGTGTGGCGGAGGCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 307
Db 649 CCCCGGGACGGGTGCGGCGGTGCGAGCGCGCGCGGAGCGGCGCGCGGAGGTG 708
QY 308 GGNCGCGGAGGTCACGACGAGCCATATACGGGCGGCGCCAGCCCCCTACTGCT 367
Db 709 GTACGGCGCGGCGACGCGGCGGGGCGCTAAGAAATCCGCTGTGTAAGCGCCCTACTCGT 768
QY 368 ACATCGCGCTCATCGGCATGGCCATCCGCGACTCGGCGGCGGCGGCTTGACGCTGGGCG 427
Db 769 ACATCGCGCTCATCACCATGGCCATCTCGAGAGCCCCAAGAGCGCTGACGCTCAGCG 828
QY 428 AGATCAACAGTACCTCATGCGGCAAGTTCCTCTTTCGCGGCGAGCTACAGGCGTGGC 487
Db 829 AGATCTGCGAGTTTCATCAGCAGCGGCTTCTTACTACCGGAGAGTTTCCCGCTTGGC 888
QY 488 GCAACTCCGTGGCGCAACCTTTCCTCAACGACTGCTTCTCAAGGTGCTGCGCGGAC 547
Db 889 AGNACAGCATCGTCAACACTGTCGCTCAACGACTGCTTCTCAAGTTCGCGCGGAC 948
QY 548 CTTGCGGCGCTGGGCAAGCAACTACTGAGTCTCAACCCCAACGAGTACACT 607
Db 949 ---CGGGCAACCGGGCAAGGCACTACTGAGCGCTCGACCGGAGTTCGCGAGATATGT 1005
QY 608 TGGCGAGCGGCTTTCGCGCGCGCGCGCGGAGGCGCTCAGCCACGCGGCGGCTCCCGG 667
Db 1006 TCGAACAGCGAGCTTCTGCGGCGCGCGCAAGCGCTTCA---AGCGCAGCGGCTACTCG 1062
QY 668 CGCGCGGCTGGGCGC 683
Db 1063 ATCCGCGCTGGGAC 1078

ORGANISM: Homo sapiens
US-09-220-132-21
Query Match 15.3%; Score 184.4; DB 4; Length 1965;
Best Local Similarity 54.4%; Pred. No. 1.2e-21;
Matches 492; Conservative 0; Mismatches 386; Indels 26; Gaps 5;
QY 278 GGGCGGCGGCGGCGCGCGCGCGGAGCGGCGGAGCGGCGGAGCGGAGCGGCGGCGGCG 337
Db 116 GCGGCGGCGGCGGCGCGGCGCGGAGCGGCGGCGGAGCGGCGGCGGAGCGGCGGCGGCG 175
QY 338 ATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 397
Db 176 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 235
QY 398 ACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 457
Db 236 GCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 295
QY 458 CTTTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 517
Db 296 CTTTCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 355
QY 518 ACAGCTCTTCGTAAGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577
Db 356 ACAGGTGCTTCAATCAAGCTGCTTAAGGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 412
QY 578 GGATGTCTAACCCCAACAGCGAGTACACTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 637
Db 413 GGACCATCGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 472
QY 638 AGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 694
Db 473 GCGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 532
QY 695 CGGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 754
Db 533 TGGGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 592
QY 755 GCGCGCG-----CGCGCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 808
Db 593 CGCTCGGCTGCGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 652
QY 809 TCGCATTCGACAGCATCTCGCGCAAGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 868
Db 653 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 712
QY 869 CGGCGGAGCGGCTTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 922
Db 713 CGCATATGTCGCGCAACCGCGGCTTCACCTTACATGCGCAGCTGCGCGGCGGCGGCGGCGGCGG 772
QY 923 CGCGGCTTCCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 982
Db 773 CGGCGGCGGCTCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 832
QY 983 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1034
Db 833 GCAGCCCGGTACCTCGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 892
QY 1035 GCGGCTCTCTGTCACCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1094
Db 893 CGAGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 952
QY 1095 GCGCG 1154
Db 953 CCCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1012
QY 1155 AGTC 1158
Db 1013 ACTC 1016

RESULT 5

Query Match 15.5%; Score 187.6; DB 5; Length 1860;
Best Local Similarity 58.3%; Pred. No. 3.6e-22;
Matches 394; Conservative 0; Mismatches 264; Indels 18; Gaps 3;
QY 8 TGGAGGTGTTGTCCTCGCGGCGCCACCGGGAGACAGCAGGCGAGTACCTGGAGGGCG 67
Db 421 TCGACGTGTTGGGGAGGCGAGGACGACGAGGAGGAGGACGATGACGACGAGGGCG 480
QY 68 CGGGCGGACGACGCGCTCCCGCTGTCGGCGCGGAGACGACTCCCTGGGCTCAG 127
Db 481 GCGGCGGCGCGCGCGGGGGTCCCGTTCGCGAGCTCGGCCACGCGGCGAGGCGT 540
QY 128 ATGGGAGTGCAGCGGCGCAAGCCGTCGCGCGGCGGCGGCCCGCCAGATACGACGAGGCGAGC 187
Db 541 CTTACCGCGGAGGTGATC-----TCGAGGACCTGGAGGAGGAGGACGACG 588
QY 188 GCGAACAGAGTGCAGGAGCGGCGCGGCGCGGAGGAGCGATCCCGCGAGCGTGTG 247
Db 589 ATGACCTGTGTGCGCCCCCGCGCGCGCTGTCGCCCGCGCTCCCGGCTCTGCGCGCG 648
QY 248 CAGCGGTGTGGCGGAGGCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 307
Db 649 CCCCGGGACGGGTGCGGCGGTGCGAGCGCGCGCGGAGCGGCGGCGGCGGAGGTG 708
QY 308 GGNCGCGGAGGTCACGACGAGCCATATACGGGCGGCGCCAGCCCCCTACTGCT 367
Db 709 GTACGGCGCGGCGACGCGGCGGGGCGCTAAGAAATCCGCTGTGTAAGCGCCCTACTCGT 768
QY 368 ACATCGCGCTCATCGGCATGGCCATCCGCGACTCGGCGGCGGCGGCTTGACGCTGGGCG 427
Db 769 ACATCGCGCTCATCACCATGGCCATCTCGAGAGCCCCAAGAGCGCTGACGCTCAGCG 828
QY 428 AGATCAACAGTACCTCATGCGGCAAGTTCCTCTTTCGCGGCGAGCTACAGGCGTGGC 487
Db 829 AGATCTGCGAGTTTCATCAGCAGCGGCTTCTTACTACCGGAGAGTTTCCCGCTTGGC 888
QY 488 GCAACTCCGTGGCGCAACCTTTCCTCAACGACTGCTTCTCAAGGTGCTGCGCGGAC 547
Db 889 AGNACAGCATCGTCAACACTGTCGCTCAACGACTGCTTCTCAAGTTCGCGCGGAC 948
QY 548 CTTGCGGCGCTGGGCAAGCAACTACTGAGTCTCAACCCCAACGAGTACACT 607
Db 949 ---CGGGCAACCGGGCAAGGCACTACTGAGCGCTCGACCGGAGTTCGCGAGATATGT 1005
QY 608 TGGCGAGCGGCTTTCGCGCGCGCGCGGAGGCGCTCAGCCACGCGGCGGCTCCCGG 667
Db 1006 TCGAACAGCGAGCTTCTGCGGCGCGCGCAAGCGCTTCA---AGCGCAGCGGCTACTCG 1062
QY 668 CGCGCGGCTGGGCGC 683
Db 1063 ATCCGCGCTGGGAC 1078

RESULT 4
US-09-220-132-21
; Sequence 21, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shvian, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE OF INVENTION: 07334-074001
; CURRENT FILING DATE: 1998-12-23
; PRIOR FILING DATE: 1998-03-25
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1965
; TYPE: DNA


```
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 475..2133
; US-083-352-1

Query Match 14.5%; Score 175.6; DB 3; Length 3946;
Best Local Similarity 50.1%; Pred. No. 2.9e-20;
Matches 552; Conservative 0; Mismatches 534; Indels 16; Gaps 4;

QY 23 CTGCGCGGCGCCAGGGGCAACAGCAGGCGAGTGACCTGAGGGCGGGCGGCGCAGGCG 82
DB 372 CTCGGCGGGCGCGGCGGACTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGG 431
QY 83 CGCGCTCCCGCTGTCGGCGCGGAGAGACGACTCCCTGGGCTCAGATGGGCGACTGCGCG 142
DB 432 GTGGGGGGCGCGGGCGCGCGGGCGCGGAGCGGGGGCCATCGAGCGC-GCTACT 490
QY 143 CCAAGCGGTCCGGGGCGGGCGCGCGAGAGATACGAGGCGGACGCGGCAAGAGATGCGG 202
DB 491 CCGTGTCCAGCCCCAACTCCCTGGGAGTGTGCTTACCTCGGCGCGGAGAGAGCTACT 550
QY 203 GAGCGGGCGGGCGGGGAGGCGGATCCCGGCGAGCTGCTGCAGCGGTGGCGG 262
DB 551 ACCGCGGGCGCGGGCGGGCGGGCGGGCGGTACACCGCATGCGCGGCCCATGAGCG 610
QY 263 AGGCGCGGAGGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGG 322
DB 611 TGTACTCGACCCCTGCGCGCGGAGCAGTACCGCGGGCGGATGGCGCGCGCTACGGGC 670
QY 323 CAGCGAGCAAGCATATACGCGCGGCGCC-----NAGCCCCCTACTGTTACATCGCG 376
DB 671 CTTACAGCGCGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
QY 377 TCATCGCATGGCCATCCGCGACTCGCGCGGGCGGGCGGCTTGAGCTGGCGGAGATCAAG 436
DB 731 TCATCACCATGGCCATCCAGAGCGCGCGGAGAGAGATCACTTGAACGCACTTACC 790
QY 437 AGTACCTCATGGCAAGTTCCCTTTTCGGGGGAGCTACAGGCGTGGCGCACTCGC 496
DB 791 AGTTTCATGAGCCGCTTCCCTTTCTACGGGACAAAGCAGGCGCTGGCAGACAGCA 850
QY 497 TGGGCGCAACCTTTTCGCTCAAGGCTGTTTCTCAAGGTGCTGCGCGACCCCTCGCGCG 556
DB 851 TCGGCGCAACCTTTCGCTCAAGGCTGTTTCTCAAGGTGCTGCGCGACGAGCAAGAGC 910
QY 557 CTTGGGCGCAAGGCAACCTACTGATGCTCAACCCCAAGCGAGTGATACCTTTCGCGGCG 616
```

RESULT 7

```
US-09-083-351-3
; Sequence 3, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E. 35,430
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: UIA-029.02
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1659 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-083-351-3

Query Match 14.2%; Score 171.6; DB 3; Length 1659;
Best Local Similarity 51.3%; Pred. No. 1.3e-19;
Matches 483; Conservative 0; Mismatches 444; Indels 15; Gaps 3;
Qy 183 CGACGGGAGAGAGTGGGGAGGCGGGCGGGCGGGAGAGCGCATCCCGCAGCAGC 242
Db 57 CGGGGGGAGCAGAGTACTACCGCGGGCGGGCGGGCGGGCGGGCGGGTACACCGC 116
Qy 243 TGCTGCAGCGGTGTGGCGGAGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 302
Db 117 CATCGCGGCCCCATGAGCGTGTACTCGCACCCCTGGCAGCGCGAGTACCGGGCGG 176
Qy 303 CGCGGGGAGCGGGAGGTGACGAGCAAGCATATACGCGGGCGGCC-----AAGCC 356
Db 177 CATGGCCCGGCTACGCGCCCTACACGCGCAGCGCGCCAGGACATGCTGAAGCC 236
Qy 357 CCCCTACTCTGATCGGCTCATCGCATGGCGCATCGGCGACTCGCGGGCGGGCGCT 416
Db 237 GCCCTATAGCTACATCGGCTCATCACCATGGCCATCCAGACGCCCCCGGACAGAAT 296
Qy 417 GACGCTGGCGAGATCAACAGTAGTACCTCATGGGCAAGTTCCTCTTTTCGCGGCGAGTA 476
Db 297 CACCTGAAGCGCATCTACAGTTTCATGAGCGGCTTCCCTTCTACCGGAGCAACA 356
Qy 477 CACGGCTGGCGGAATCTCGTGGCCACAGCTTTCCTCAACGACTGCTTCGTAAGT 536
Db 357 GCAGGGTGGCAGAACAGCATTCGCCCAACCTCTCGCTCAACGAGTGTCTGTAAGT 416
Qy 537 GCTCGGAGACCCCTCGGGGCGCTGGGCAAGAGCAACTACTGATGTCTCAACCCCAACAG 596
Db 417 GCCGGGAGCAGCAAGAGCC--GGCAGAGGCGAGTCTGAGCGTGGACCGCGGACTC 473
Qy 597 CGAGTACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGC 656
Db 474 CTACAACATGTTTCAGAACGCGAGCTTCTCGCGCGCGCGCGCGCTTCAAGAGAGGA 533
Qy 657 GCCGCTCCCGCGCGCGCTGCGCGCGCGAGGAGGCC-----CGGGCTCCCGCGCG 710
Db 534 CGCGGTGAAGACAAGAGAGAGAGAGAGCTGCACTCAAGAGAGCGCGCGCGCGCGCG 593
Qy 711 CCCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCATGCGTTCGCGCGCGCGCGCAGGA 770
Db 594 CGCGCAGCGCGCGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
Qy 771 GGAGGGGCGCAGCGCGCGCGCGCGAGTTCTCAGTCTCTTCGCTTCGATCGACAGCATCTGCG 830
Db 654 GCCGCGCGTGGCATTCAGAGCATCAAGACCGAGACGGTACGTGCGCTTCGCGCGCGCA 713
Qy 831 CAAGCCCTTCGCGCAGCGTTCGCTTCAGGGACACGCGCGCGCGCGCGCGCGCGCGCG 890
Db 714 GCCCTGTGTCG 773
Qy 891 CGCGCGCGCTTCG 950
Db 774 GAGCCCCGACAGCAGCAGCAGCAGCTGTCCAGCGGAGCAGCGCGCGCGCGCGCGCG 833
Qy 951 GCGCTGTCGCGCTTCGCGCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1010
Db 834 GTCGCGCGCGCGCTCAGCGTGGAGGTGCGGATTCGCGCGCGCGCGCGCGCGCGCG 893
Qy 1011 GGCGGAGGTGCCACCGAGCGCGCGCGCGCGCTCTCTCTCTGACACTCTCTCCGCGCGCGCGCG 1070

Db 894 CGCCCCCG 953
Qy 1071 CGCCAAGCGCACTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1112
Db 954 GGGGTGCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCTCCGCGCT 995
RESULT 8
US-09-083-352-3
Sequence 3, Application US/09083352
Patent No. 6207450
GENERAL INFORMATION:
APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,352
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1659 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-083-352-3

Query Match 14.2%; Score 171.6; DB 3; Length 1659;
Best Local Similarity 51.3%; Pred. No. 1.3e-19;
Matches 483; Conservative 0; Mismatches 444; Indels 15; Gaps 3;
Qy 183 CGACGGGAGCAGAGTGGGGAGGCGGGCGGGCGGGAGAGCGCATCCCGCAGCAGC 242
Db 57 CGGGGGGAGCAGAGTACTACCGCGGGCGGGCGGGCGGGCGGGCGGGTACACCGC 116
Qy 243 TGCTGCAGCGGTGTGGCGGAGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 302
Db 117 CATGCGGCCCCATGAGCGTGTACTCGCACCCCTGGCAGCGCGAGTACCGGGCGG 176
Qy 303 CGCGGGGAGCGGGAGGTGACGAGCAAGCATATACGCGGGCGGCC-----AAGCC 356
Db 177 CATGGCCCGGCTACGCGCCCTACACGCGGAGCGCGCGCGCGCGCGCGCGCGCG 236
Qy 357 CCCCTACTCTGATCGGCTCATCGCGCATGGCGCGCGCGCGCGCGCGCGCGCGCG 416
Db 237 GCCCTATAGCTACATCGGCTCATCACCATGCGCATCCAGAACGCGCGCGGACAGAAT 296

417	Qy	GACCTGGCGGAGATCAACGAGTACTCTATGGGCAAGTTCCCTTTTTCGGCGGAGCTA	476
297	Db	CACCTCTGAACGGCATCTACCAAGTTTCATATGAGACGGTTCCCTTTTACCGGGAACA	356
477	Qy	CACGGCTGGCGCAACTCCGTGGCGCAACACTTTTCGCTCAACGACTGCTTCGTCAAGT	536
357	Db	GCAGGGCTGGCAGACAGCATCCGGCCACACCTTCGCTCAACGAGTGTTCGTCAAGT	416
537	Qy	GCTGGCGGACCCCTTCGCGGCCCTGGGGCAAGACACTACTGATGCTCAACCCCAACAG	596
417	Db	GCGCGCGACGACAAAGAACCC---GGGCAAGGGCAGCTACTGACCTCGACCCGCACTC	473
597	Qy	CGAGTACACTTTCGCGGACGGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACCGCGC	656
474	Db	CTACACATGTTTCGAGAACGGCAGCTTCTTCGCGCGCGCGCGGTTCAGAGACAGGA	533
657	Qy	GCCGCTCCCGCGCCCGGGCTTCGCGCCCGAGAGGGCCC-----CGGGCTCTCCCGCGC	710
534	Db	CGCGGTGAAGACAAGGAGGAGAGGACAGGCTGCACCTCAAGGACGCCGCCGCCCGG	593
711	Qy	CCGCGCGCGCGCGCGCGCGCGCTCGCCCCGATGCGTTCGCGCGCGCGCGCGCAGGA	770
594	Db	CGCGCAGCCCCCGCGCGCGCGCGAGCGCGACGCGCGCGCGCGTTCGCGACCC	653
771	Qy	GGAGCGCGCACGCCCGCGGGCAAGTTCTCCAGTCTCTTCGCCATCGACAGCATCTCTGC	830
654	Db	GCGGCCGTGGCATCCAGGACATCAAGACCGAGAACGGTACGTGCCCTCGCGCCCA	713
831	Qy	CAAGCCCTTCGCGACCGCTTCGCTCAGGGACAAGGCCCGCGGACAGACGCTTCAGTGGG	890
714	Db	GCCCCGTGTCGCGCGCGCGCTTGGGACGCGGACGCGCGCGCGGTGCCAAGATCGA	773
891	Qy	CGCGCGCGCTCGCGCGCGCTGCCCGGTTCCCGCGCTCTCTCCCGCGCGCGCTCGAG	950
774	Db	GAGCCCCGACAGCAGCAGCAGCTGTCCAGGGGAGAGACCCCGCGGCGAGCTGCC	833
951	Qy	GGCCCTGTTCGCGTCTTCGGCGTACGGCGCGGGGAGCGCGCGCGCTGGGCGCGCGGA	1010
834	Db	GTTCGGCGGCGCGCTCAGCTTGGAGCGGTGGGGATTTCGCGCGCGCGCGCGCGCTTC	893
1011	Qy	GGCCGAGTGCACCGACGCGCGCCCTCTCTGCTTGCACTCTCCCGCGCGCGCGCCC	1070
894	Db	GCCCCGCGCGCATAGCATAGGAGGTTTCAGCGTGGACACATCATGACTGCTGCTGG	953
1071	Qy	CGCCAAAGCCACTTCGAGGCCCGCGCGCGCGCGACCT	1112
954	Db	GGGCTCGCGCAGAGCGCGCGCGCGAGCTCAGCTCGGCT	995

RESULT 9

RES001.3
 US-07-882-292-1
 ; Sequence 1, Application US/07882292
 ; Patent No. 5324638
 ; GENERAL INFORMATION:
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Lai, Eseng
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
 ; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller
 ; STREET: Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.24
 ; CURRENT APPLICATION DATA:

1154	GACGACCC---GGGCAAGGGCACTACTGATCTCGA	CCCTCGAGCGACGAGTGTTCT	1210
Db			
610	GCCGACGGGGTCTTTCCGCGCGCGCAAGCGCTCAGCC	CACCGCGCGCGCTCCCGCGC	669
QY			
1211	ATCGCGSCACGACCGGCAAGCTGCGGCGCGCTCCAC	CACGCTCTCGGGCCAAAGTAGCC	1270
Db			
670	CCCGGGCTGCGGCGCGGAGAGGCCCGGGGCTCCCG	CGCCCGCGCGCGCGCGCGCGCC	729
QY			
1271	TTTAGCGCCGGGACAGGGCTCAGCTCCACCGGGCTC	ACCTTTCATGGACCGCG---CCGCGC	1327
Db			
730	GCCCGGGCTCGCCCGCGATGCGCTGCGCGCGCGCG	CAGGAGCGCGCGACCGCGCGC	789
QY			
1328	TCCCTCTACTGGCCCATGTGCGCCCTTCCTGCTGAC	CCACTCTGCGCCAGAGACT	1387
Db			
790	GGCAAGTTCTCCAGCTCCTTTGGCCATGACAGCATCT	CGCAAGCCCTTCGCG	843
QY			
1388	TTGAGTTTAAACGGGAGCCACTGGCTGGCTACCCAG	AGCACTCCATGCGCTTACGC	1441
Db			

RESULT 10
 US-08-331-644-1
 ; Sequence 1, Application US/08331644
 ; Patent No. 5976872
 ; GENERAL INFORMATION:
 ; APPLICANT: Tao, Wufan
 ; APPLICANT: Lai, Eseng
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/331.644

```

; OTHER INFORMATION: homology"
;
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 1883..1885
; OTHER INFORMATION: /note= "translation termination
; OTHER INFORMATION: codon"
;
US-08-331-644-1

Query Match      12.6%; Score 152.8; DB 2; Length 2830;
Best Local Similarity 52.8%; Pred. No. 1.3e-16;
Matches 409; Conservative

Qy 82 GCGCCCTCCCGCTGTCGGCGCGGAGACGACTCCCTCGGTTCAGATGGGACTCGCG 141
Db 674 GCCCGCAGCCCCGCGAGCGCGCGCCGCCAGCAGCGACGACGACGAGGCCCCAG 733
Qy 142 GCCAAGCGTCCCGGGCGCGCGCCAGAGATACCGAGGGCGACGGCAACAGAGTGG 201
Db 734 CCCTTTCTGCTCCCGCGCTCCGCCGCCCTGGAGCGGGGCCAAGGCTCAGCGACT 793
Qy 202 GGAGGGCGGGCGGGCGCGGAGGAGGCGCATCCCGGCAGCAGCTGCTCAGCGGTGG 261
Db 794 AAGAGGCGAGCGCGCGGGCTGCGGAGCTGGCCCGTCGGCGCGGACGAGAGGAG 853
Qy 262 GAGGGCGGAGCGCGGGCGCGGGCGCAGGGCGCGGGCGCGGGGAGCGCGGAGGGT 321
Db 854 AAGGGCGCGGGCGCTCGGGGGGAGGAGAGAGGGGGCGGGCGAGGGCGGCAAGG 913
Qy 322 GCACGCGCAAGCCATATACGGCGCGGGCC-----AAGCCCCCTACTCGTAC 369
Db 914 GAGGGGGCAGGAGGGCGNACAGAACACGGCAAGTACGAGAAAGCGCGCTTCACTAC 973
Qy 370 ATCGCGCTATCGCCATGCGCATTCGCGACTCGCGGGCGGGCGGCTTGACGTGGCGGAG 429
Db 974 AACCGCTCATCATGATGGCATCAGGCAGAGTCCGCGAAGAGGCGCTGACGCTCAACGGC 1033
Qy 430 ATCAACGAGTACCTCATGGCGAGTTCCTCTTTTTCGGCGGAGCTACACGGGCTGGCGC 489
Db 1034 ATTACGAGTTTCATCATGAAGACTTCCCTTACTACCGAGAGAACAGAGGGCTGGCAG 1093
Qy 490 AACTTCGTGGCCACAACTTTGCTCAACGACTGCTTGTCTAAGGTGCTGCGGACCCCC 549
Db 1094 AACTCCATCGCCACAACTGTCCCTCAACAAGTGCTTGTGAAGGTACCGCGCCACTAC 1153
Qy 550 TCGGGCCCTGGGGCAAGGACACTACTGATGCTCAACCCCAACGAGGATACACTTC 609
Db 1154 GACGACCC---GGGCAAGGGCACTACTGATCTGGACCCCGTCGAGCGACAGTGTTTC 1210
Qy 610 GCCGACGGGGTCTTCGCGCGCCCGCGCAAGCGGCTCAGCACCGCGCGCGGTCCCGCG 669
Db 1211 ATCGGGCGCAGACCGGCAAGCTGCGGCGCGCTCCACACAGTCTCGGGCCAAAGCTAGCC 1270
Qy 670 CCGGGGTGGGCGCCGAGGAGGCCCGGGCTCCCGCGCGCCCGCGCGCGCGCGCGCC 729
Db 1271 TTAAAGCGCGGGCAGGGCTCAGCTCCACGGGCTCACCCTTATGAGACCGCG---CGGC 1327
Qy 730 GCCCGGCTTCGCCCGCATGCGCTCGCGCGCCCGCAGGAGGAGGCGCGGACCCCGCG 789
Db 1328 TCCCTCTACTGGGCCATGTGCCCTTCCTGTCCCTGCAACCACTTCGGGCCAGCAGCACT 1387
Qy 790 GGCAAGTTCTCAGTCTCTTGGCATCGACAGCATCTCTGGGCAAGCCTTCCGC 843
Db 1388 TTGAGTTTACCAACGGGACCACTCGGGCTATCCCGAGCACCCTATGCTTACAGC 1441

```

```

RESULT 11
PCT-US93-04102-1
; Sequence 1, Application PC/TUS9304102
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND US93 THEREOF

```

NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: John P. White
 STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/04102
 FILING DATE: 19930430
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/882,292
 FILING DATE: 13-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 41472A-PCT/
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-977-9550
 TELEFAX: 212-664-0525
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2830 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: N
 ANTI-SENSE: N
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 443..1882
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 926..1255
 OTHER INFORMATION: /note= "nucleotide sequence encoding DNA
 OTHER INFORMATION: binding domain homology"
 FEATURE:
 NAME/KEY: misc_signal
 LOCATION: 1883..1895
 OTHER INFORMATION: /note= "translation termination codon"
 PCT-US93-04102-1

Query Match 12.6%; Score 152.8; DB 5; Length 2830;
 Best Local Similarity 52.8%; Pred. No. 1.3e-16;
 Matches 409; Conservative 0; Mismatches 347; Indels 18; Gaps 3;
 QY 82 GCGCCCTCCCGTTCGCGGGGGGAGAGAGACTCCCTGGGCTCAGATGGGACTGCCGG 141
 DB 674 GCGCCCGAGCCCCGCGAGGCGCGCGGCCCGCCAGCAGCGGACGACAAAGGGGCCCGCCAG 733
 QY 142 GCCAAGCCGTCGCGGGCGCGCGCCAGAGATACGCGGGGCGACGGCGAAACAGAGTGCG 201
 DB 734 CGCTTCTGCTCCCGCGCTCCCGCGCTCGAGCGGGCCAAAGCTGACGCATTTGAGGCC 793
 QY 202 GGAGGCGGGCGGGCGCGAGGAGGCGATCCCGGACAGAGCTGCTGACGCGTGTGTGGCG 261
 DB 794 AAAGGCGAGCAGCGCGGGCGCTGCGGAGCTGGCGCGCTCGGGCGGACGAGAAGAGAG 853
 QY 262 GAGGGCGCGAGCGCGGGCGGGCGAGGCGCGGGCGGGCGGGGAGCGCGGAGCGCGAGG 321
 DB 854 AAGGGCGCGGCGCTGGGGGGGAGGAGAAAGAGGGGGCGGGCGAGGGCGGCGAGGACG 913
 QY 322 GCACGCAAGCAAGCCATATACGCGCGCGCGCC-----AAGCCCCCTTACTCGTAC 369

DB 914 GAGGGGGCGAGGAGGGCGCAAGAAACAACGCGCAAGTACGAGAGCGCGCTTCCACCTAC 973
 QY 370 ATCGCGCTCATCGCATGGCGACTCCGCGACTCGGCGGGCGGGCGCTTGAAGCTGGCGGAG 429
 DB 974 AACGGCTCATGATGGCCATCAGCAGAGTCCCGAGAGCGCTTACGCTCAACGGC 1033
 QY 430 ATCAACGAGTACCTCATGGGCAAGTTCCCTTTTTCGCGGCGAGCTACAGGGTGGCGC 489
 DB 1034 ATCTACGAGTTTCATCAAGAACTTCCCTTACTACCGCGAGAAACAAGCAGGGGTGGCAG 1093
 QY 490 AACTCCGTGGCCCAACCTTTCCTCAAGACTGCTTCTCAAGGTGCTGCGCGACCCC 549
 DB 1094 AACTCCATCCGCAACACTGCTCCCTCAAGAGTCTTCTGAGGTACCGCGCCTAC 1153
 QY 550 TCGCGGCTGGGGCAAGGCAACTACTGATGCTCAACCCCAACAGCGAGTACACCTTC 609
 DB 1154 GACGACCC--GGGCAAGGGCACTACTGATGTGTGACCCGCTCGAGCGAGCGTGTTC 1210
 QY 610 GCCGACGGGGTCTTCGCGCGCGCGCAAGCGCTCAGCCACGCGCGCGGTCCCGCG 669
 DB 1211 ATCGCGGCGAGCCGCGAAGCTCGGGGCGCTCCACAGGTCTCGGGCCAAGCTAGCC 1270
 QY 670 CCCGGCTGGCGCCGAGAGGGCCCCGGGCTTCCCGCGCGCGCGCGCGCGCGCGCGCC 729
 DB 1271 TTAAAGCGCGGGCAGCGCTCACCTCCACGGGCTCACCTTCATGAGACCGCG--CGGCG 1327
 QY 730 GCCCGGCGCTCGCGCGCATGCGCTCGCGCGCGCGCGCGCGAGGAGCGCGCGCGCGCG 789
 DB 1328 TCCCTCTACTGGCCCATGTGCGCCCTTCTGTCTTGCACACCCCTCGCGCGCAGGACT 1387
 QY 790 GCGAAGTTCTCCAGCTCCTTCGCCATCGACAGCATCTCGCGCAAGCCCTTCCCG 843
 DB 1388 TTGAGTTACAAGGGGACCACTCGGCGCTACCCAGCCACCCCATGCGCTACAGC 1441

RESULT 12

US-09-976-594-927
 ; Sequence 927, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 927
 ; LENGTH: 1309
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 3687719CB1
 US-09-976-594-927

Query Match 11.1%; Score 134.6; DB 4; Length 1309;
 Best Local Similarity 57.6%; Pred. No. 1e-13;
 Matches 261; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
 QY 340 ACGCGGGCGCCCAAGCCCGCTTACTCGTACATCGCGCTCTCGGCATGGGCATCCGGAC 399
 DB 93 ACTGAGCAACCAAGCCCTCCCTACAGCTACATCGCCCTTATGCTATGGGCATCCAGAGC 152
 QY 400 TCGCGCGGGCGCGCTTTCGCTGGCGGAGTCAACAGTACCTTCATGGGCAAGTTCCCG 459
 DB 153 TCACGGGGGCGAGCGGGCGACCCCTCAGTGGCATCTACCGCTACATCATGGCGCGATTGCC 212
 QY 460 TTTTTCGCGCGAGCTACACGGGCTGGCGCAACTCGCTGCGCCACAACTTTTCGCTCAAC 519

Db 213 TTCTACCGCCACACCGCGCGGCTGGCAGAACAGATCCGCCAATCTGTCACTCAAC 272
Qy 520 GACTGCTTCGTCAGTGTGCGGACCCCTCGCGGCGCTGGGGCAAGGACAACTACTGG 579
Db 273 GAGTGTCTTGTCAAGTGTGCGCGCGGATGACCGCAAGCC---AGGCAAGGCGAGTACTGG 329
Qy 580 ATGCTCAACCCACACGAGTACACCTTCGCGGAGCGGGTCTTCGCGCGCGCGCGCAAG 639
Db 330 AGCTGACCTGACTGCTCCAGCATGTTTGAGCAGCGAGCTTCTACGCGCGCGCGCGC 389
Qy 640 CGCTTACGACCGCGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699
Db 390 CGCTTACCGCGCGCGCGCGCGGCTGAGGCGACCGCGCGCGCGCGCGCGCGCGCG 449
Qy 700 CTCCCG 759
Db 450 CCCTTACG 790
Qy 760 GCG 792
Db 510 TCATTCACCGAGGCTGCCAGATCCCAAGGCG 542

RESULT 13

US-08-331-644-4
; Sequence 4, Application US/08331644
; Patent No. 5975872
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,644
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472A-PCT-US
; TELEPHONE: 212-278-0400
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-331-644-4

Query Match 10.8%; Score 130.2; DB 2; Length 1155;
Best Local Similarity 67.5%; Pred. No. 5.1e-13;
Matches 199; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

Qy 352 AAGCCCCCTACTCGTATACGCGCTCATCGCATGSCCATCCCGACTCGCGCGCGGG 411
Db 499 AAGCGCGCTACAGCTATATAGCGCTCATCACCATGGCGATCCAGAACGCGCCAGAGAAG 558
Qy 412 CGCTTACGCTGCGGGGAGATCAACGAGTACCTCATGGGCAAGTTCCTTTTCCGCGGC 471
Db 559 AAGATCACTCTGAACGGCATCTACCGATTCACTGAGACGCTTTCCCTCTTACCGGAG 618
Qy 472 AGCTACAGCGGCTGGCGCACTCGTGGCGGCAACCTTTTCGCTCAAGACTGCTTCGTC 531
Db 619 AACAGCAGGGCTGGCAGACAGCATCGGCCAACCTGTCTCAATGAGTGTCTGTG 678
Qy 532 AAGTGTCTGCGCGACCCCTCGCGGCTTGGGGCAAGGACAACTACTGATCTCAACCCC 591
Db 679 AAGTGTGCGCGGAGCAAGAACGCG---GGGCAAGGCGAGCTACTGAGCTCGACCG 735
Qy 592 AACAGCAGTACACTTCGCGGAGCGGGTCTTCGCGCGCGCGCGCGCGCGCTCA 646
Db 736 GACTCTACCAACATGTTGAGATGGCAGCTTCTCGCGCGCGCGCGCGCGCTTCA 790

RESULT 14

PCT-US93-04102-4
; Sequence 4, Application PC/TUS9304102
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04102
; FILING DATE: 19930430
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-977-9550
; TELEFAX: 212-664-0525
; TELEX: 42523 COOP UI
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US93-04102-4

Query Match 10.8%; Score 130.2; DB 5; Length 1155;
Best Local Similarity 67.5%; Pred. No. 5.1e-13;
Matches 199; Conservative 0; Mismatches 93; Indels 3; Gaps 1;

Qy 352 AAGCCCCCTACTCGTATACGCGCTCATCGCATGSCCATCCCGACTCGCGCGCGGG 411
Db 499 AAGCGCGCTACAGCTATATAGCGCTCATCACCATGGCGATCCAGAACGCGCCAGAGAAG 558

QY 412 CGCTTGACGCTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTCGCGGC 471
Db 559 AAGATCCTCTGAACGGCACTACCGATTCATGAGACCGTTCCCTTCTACCGCGAG 618
QY 472 AGCTACACGGGTGGCGCAACTCCGTGCGCCAAACCTTTGCTCAACGAGTGTCTTC 531
Db 619 AACAGCAGGGTGGCGAAGCAGCATCCGCCAACCTGTCACTCAATGAGTGTCTCGTG 678
QY 532 AAGGTGCTCGCGACCCCTCGGGCCCTGGGGCAAGGCAACTACTGATGCTCAACCCC 591
Db 679 AAGTCCCGCGAGCAGCAAGAGCC--GGCAAGGCGAGCTACTGGACGCTGACCCG 735
QY 592 AACAGCGAGTACACCTTCCGCGACGGGTCTTCGCGCGCGCGCAAGCGCTCA 646
Db 736 GACTCCTCAACATGTTTCGAGATGCGAGCTTCTTCGCGCGCGCGCGCTTCA 790

RESULT 15
US-09-976-594-495
; Sequence 495, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 495
; LENGTH: 5080
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 206866.1
; NAME/KEY: unsure
; LOCATION: 2468
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-495

Query Match 9.7%; Score 117.6; DB 4; Length 5080;
Best Local Similarity 65.4%; Pred. No. 5.3e-11;
Matches 204; Conservative 0; Mismatches 104; Indels 4; Gaps 2;
QY 324 ACGCAGCAAGCCATATACGCGGCGCCCAAGCCCTACTCGTACATCGGCTCATCGC 383
Db 22 ACGTCCAGCGCAGCTACCCGCGCGCCAGCGCGCTACTCGTACATCTCGCTCATCAC 81
QY 384 CATGGCCATCCCGACTCGCGGCGG-GCGCTTGACGCTGGCGGAGATCAACGAGTACC 442
Db 82 CATGGCCATCCAGCAGCGCCCGCCAGCAGATGCTCACGCTGAGCGAGATCTACCAGTGA 141
QY 443 TCATGGCAAGTTCCTCTTTTCGCGGCGAGTACACGGGTGGCGCAACTCCGTGCGCC 502
Db 142 TCATGACCTCTTCCCTATTACCGGCAACACAGCGCTGGCGAGACTCCATCCGCC 201
QY 503 ACAACCTTTGCTCAACGACTCTTCTGTCAGGTGCTGCGGACCCCTCGCGGCGCTGG 562
Db 202 ACTCGCTGTCTCAATGACTGCTTCTGTCAGGTGGCAGCTCCCGCGCAAGCC--GG 258
QY 563 GCAGGACACTACTGGATGCTCAACCCCAAGCGAGTACACCTTCCGCGCGGGTCT 622
Db 259 GCAAGGGCTCTTACTGAGCGCTGCACCCGGACTCCGCGCAACATGTTTCGAGAACGCGTCT 318
QY 623 TCCGCGCGCGCC 634
Db 319 ACTTGGCGCGCC 330

APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula
APPLICANT: Gorbacheva, Bella
APPLICANT: Hoersch, Sebastian
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Monsey, Angela M.
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumel
APPLICANT: Anderson, Dustin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/307,982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314,356
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/325,020
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 135
LENGTH: 2487
TYPE: DNA
ORGANISM: Homo sapiens
US-10-205-823-135

Query Match 16.7%; Score 201.4; DB 14; Length 2487;
Best Local Similarity 52.1%; Pred. No. 2e-33;
Matches 574; Conservative 0; Mismatches 512; Indels 16; Gaps 5;
QY 83 CGCGTCCCGGTGTCGGCGCGGAGACGACTCCCTGGGCTCAGATGGGACTGCGCGG 142
DB 517 CGACGACGAGGCGCGGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 576
QY 143 CC--NAGCGCTCGCGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
DB 577 GCGCGCTCGTACGCGCGGAGGACGAGCTGAGGATCTGAGGAGGAGGAGGACGACGA 636
QY 201 GGGAGCGCGCGCGCGCGCGGAGGCGGATCCCGGCGAGCGTCTGCGAGCGGTGGTGGC 260
DB 637 TGACATCTGTGGCGCGCGCTGTGGGGTCTCCCGCGCGCGCGCGCGCGCGCGCGCG 696
QY 261 GGAGGCGCGGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
DB 697 GCGGCGCGGAGGCGCGGTGGGCGCGG---CGGCGGCGCGCGCGCGCGCGCGCGCGGAG 753
QY 321 TGCAGCGAGCAAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 380
DB 754 CGCGGTAGCGT 813
QY 381 CGCCATGCGCATTCGCGACTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 440
DB 814 CACTATGGCCATCTCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 873
QY 441 CTTATGGGCAAGTTCCCTTTTCCGCGGAGTACACGCGGCTGGCGCACTCCGTCGCG 500
DB 874 CATACGCGCGCTTCCCTTACTACCGGAGAGTTCCCGCGCGCGCGCGCGCGCGCGCG 933
QY 501 CCACAACCTTTGCTCAACGACTGCTGCTCAAGGTGTCGCGCGCGCGCGCGCGCGCGCG 560
DB 934 CCACAACCTCTCGCTCAACGACTGCTGCTCAAGATCCCGCGCGCGCGCGCGCGCG 990
QY 561 GGGGAGGACAACTACTGATGCTCAACCCCAAGCGAGTACACCTTCGCGCGCGCGGT 620

DB 991 GGGCAAGGGCACTACTTGACCGTGGACCCGCGAGTCCGCGGACATGTTGACACAGCGAG 1050
QY 621 CTTCCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGCGCGCGTCCCGCGCGCGCGCGT 680
DB 1051 CTTCTCTGCGCGGAGGAAGCGTTCAAGCGGACGCGCTGCTCCACCAACCGCGCGCG 1110
QY 681 GCGCGAGGAGGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
DB 1111 CGCGGAGTCTGTGCTGCTGCGGCGCGGAGCGCGAGGCGGCGCGCGCGCGCGCG 1170
QY 741 GCGCGCATGCGCTCG 800
DB 1171 CGCGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTA 1224
QY 801 CAGCTCTTTCGCGCATCGACAGCATCTGCGCAAGCGCTTCCGAGCGCGTGGCTCAGGGA 860
DB 1225 CGCGCGCTTACGCGTGGCGTACGCGCTGCGCGCTTACGCGCGCGCGCGCGCGCG 1284
QY 861 CAGCGCGCGCGCGCGCGCGCTTCAAGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 920
DB 1285 CTTTCG 1344
QY 921 CCGCGCGCTCTCCCG 980
DB 1345 GCG 1404
QY 981 GGGCGAGCG 1040
DB 1405 GCACCGCGCT--CGGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 1462
QY 1041 CTTGCTTGCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100
DB 1463 GCG 1522
QY 1101 GCG 1160
DB 1523 AGTTTGGCG 1582
QY 1161 NGTCTTGGCG 1182
DB 1583 TGGCGCTCGCGCTGCGCGGTG 1604

RESULT 7

US-10-101-510-133
; Sequence 133, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 133
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-133

Query Match 16.0%; Score 193.4; DB 14; Length 2271;
Best Local Similarity 52.8%; Pred. No. 9.6e-32;
Matches 490; Conservative 0; Mismatches 427; Indels 11; Gaps 3;
QY 255 GGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
DB 440 GCGCGCGCGCGCGCGGCGGAGCGGTGGGCGCGCGCGCGCGCGCGCGCGCGCG 499
QY 315 CGAGGCTGCGCGAGCGAGCGCATATACGCGGCGCGCGCGCGCGCGCGCGCGCG 374

500 CGGAGCGGGTGGCGGCGCCAGAACCGCTGGTGAAGCGCCCTACTCGTATATCGC 559
375 GCTATCGCCATGCGCATTCGCGACTCGCGGCGCGGCGCTTACCGCTGGCGGAGATCAA 434
560 GCTCATCACTATGCGCATTCGCGAGAGCGCCCAAGAGAGCGCTGACGCTGAGCGAGATCG 619
435 CGAGTACCTCATGGGCAAGTTCCTCTTTTCGCGCGCAGCTACACGGGCTGGCGCAATC 494
620 TGAGTTTCATCAGCGCGCGCTTCCCTACTACCGGAGAGTTCCTCGCTGCGAGNACAG 579
495 CGTGCGCCACAACCTTTTCGCTTCAAGACTGCTTTCGCTCAAGGCTGCGCGACCTTCGCG 554
680 CATCGCGCACAACTCTTCGCTCAAGACTGCTTTCGCTCAAGTTCCTCGCGGAGCGCC---GG 736
555 GCCTTGGGCAAGAGAACTTCTGCTGCTCAAGCGGAGAGTTCCTCGCTGCGCGA 614
737 CAACCGCGGCAAGGCACTTCTGAGAGTTCGAGAGCGGAGTTCGCGCGCATGTTGACAA 796
615 CGGCGTCTTTCGCGCGCGCGCTTCAAGCGGAGTTCGAGAGCGGAGTTCGCGCGCGCG 674
797 CGGAGCTTCTGCGCGGAGGAGGCTTCAAGCGGAGCGGAGTTCGCGCGCGCGCGCG 856
675 GCTGGGCGCGGAGGCGCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 734
857 CGCGCGCGCGGAGTCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 916
735 GGCTTGGCGCGGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 794
917 GCGAGCGCGCGCGCTTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
795 GTTCTCAGCTCTTTCGCGCATGAGAGATCTGCGCGAGCGCGCGCGCGCGCGCGCGCG 854
971 CGGCTACGGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 1030
855 CAGGAGACGCGCGCGCGCGCGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 914
1031 GGCGCTTTCG 1090
915 CGGTTTCGCGCGCTTCTCG 974
1091 CG 1150
975 CG 1034
1151 CG 1208
1035 GCGCTTCTGCGCGCTTCTCG 1094
1209 GCG 1268
1095 GCG 1154
1269 GCG 1328
1155 AGTCCGCGCGCTTTCG 1182
1329 CAGGCGCTGCGCGCTTTCG 1356

RESULT 8
US-09-960-706-947
; Sequence 947, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Mungen, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; TITLE OF INVENTION: Gene Expression Profiles
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07

; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 947
; LENGTH: 2187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 U13220
US-09-960-706-947

Query Match 15.3%; Score 184.4; DB 10; Length 2187;
Best Local Similarity 54.4%; Pred. No. 7.3e-30;
Matches 492; Conservative 0; Mismatches 386; Indels 26; Gaps 5;

QY 278 GGGCGGCGGCG 337
DB 338 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 397
QY 338 ATACGCGCGGCG 397
DB 398 TGGCGGCGGCGCGGCG 457
QY 398 ACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 457
DB 458 GCTCG 517
QY 458 CTTTTCG 517
DB 518 CTTTTCG 577
QY 518 ACGACTGCTTTCG 577
DB 578 ACGAGTCTTTCATCAAGCTGCTTAAAGGCGCTTAAAGGCGCTTAAAGGCGCTTAAAG 634
QY 578 GGATGCTCAACCCCAACAGCGAGTACACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCG 637
DB 635 GGAACATCGACCG 694
QY 638 AGCGCTCAGCG 694
DB 695 GCGGCTTTCAGCGGGAAGTGCAGCGCGCTCAAGCGCGCTCAAGCGCGCTCAAGCGCG 754
QY 695 GCGGCTTTCAGCGGGAAGTGCAGCGCGCTCAAGCGCGCTCAAGCGCGCTCAAGCGCG 754
DB 755 TGGGCTTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
QY 755 GCG 814
DB 815 CGCTCGGCTGCGCAAGCG 874
QY 809 TCGGCATCGACAGCATCTTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 868
DB 875 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 934
QY 869 CCGGGAAGAGCGCTTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 922
DB 935 CGCACATGTGCGCCCAACCGCGGTTCCACCTACATGCGCGCGCGCGCGCGCGCGCGCG 994
QY 923 CCGCGCTTCTTCG 982
DB 995 CCGGCGGCGCTGCGTGGCG 1054
QY 983 CGGAGCG 1034
DB 1055 GCAGCGCGGTACCTTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1114
QY 1035 GCGCGCTTCTGCTGCG 1094
DB 1115 CGAGCGCTTTCG 1174
QY 1095 GCGCGGCGGCG 1154


```

Db      1175  CCTGTAGCGCCGAGCAGCAACCCCGCCGCTCGGCGAGGCGCTGCACTCCAGCATGTCTCTCT 1234
Qy      1155  AGTC 1158
Db      1235  ACTC 1238

RESULT 9
US-09-292-862-1
; Sequence, 1, Application US/09292862
; Publication No. US20030013087A1
; GENERAL INFORMATION:
; APPLICANT: Walter, Michael A.
; APPLICANT: Jordan, Tim
; APPLICANT: Raymond, Vincent
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE PFEAC3 GENE FOR
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT
; TITLE OF INVENTION: DYSGENESIS
; FILE REFERENCE: 07540/020003
; CURRENT APPLICATION NUMBER: US/09/292,862
; CURRENT FILING DATE: 1995-04-16
; EARLIER APPLICATION NUMBER: 60/084,784
; EARLIER FILING DATE: 1998-05-08
; EARLIER APPLICATION NUMBER: 60/082,206
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-292-862-1

```

Query Match	13.4%;	Score 161.6;	DB 10;	Length 1662;	
Best Local Similarity	51.3%;	Pred. No. 4.5e-25;			
Matches 483;	Conservative 0;	Mismatches 444;	Indels 15;	Gaps 4;	
Qy	183	CGACGGCGAACAGAGTGC	CGGAGACGCGGCGGCGGCGGAGAGCGCATCCGGCAGCAGC	242	
Db	57	CGGCGGCGAGCAGAGCTA	TACCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTACACCGC	116	
Qy	243	TGCTGCAGCGGTGTGTG	CGGAGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	302	
Db	117	CATGCCGCGCCCATGATG	AGCGTGTA	CTGCACCTTCGCACGCCGAGCAGTACCCGGGCGG	176
Qy	303	CGCGGGGAGCGGCGAGGG	TGCACGCACGAGCCATATACGCGGCGGCGCC-----AAGCC	356	
Db	177	CATGGCCGCGGCCCTAC	GCGGCCCTACACGCGCGCAGCCGACGCCAACGACATGTTGTAAGCC	236	
Qy	357	CCCTACTCTGTACATCG	CGCTCATCGGCATCGGCATCGGCATCGCGACTCGCGCGGGCGGCGCTT	416	
Db	237	GCCTCTA	TAGCTACATCGCGGTCA	TACCATGCGCCATCCAGAACGCCCGCGACAAAGAT	296
Qy	417	GAGCCTGGCGGAGATCA	ACGAGTACCTTCATGGGCAAGTTC	CCCTTTTTCGCGCGCAGCTA	476
Db	297	CACCTGAA	CGGCATCTAC	CAGTTTCATATGACCGGCTTCCCTTTCTACCGGACACAA	356
Qy	477	CAGGGGTGGCGAACTC	CGTGGGCCAACACCTTTTGCTCAACGA	CTGCTTCGTCAAGGT	536
Db	357	GCAGGGCTGGCAGAA	CAGCATCGCCCAACCTCTCGCTCAACGAGTGTCTCGTCAAGGT	416	
Qy	537	GCTGCGGACCCCTCG	GCGCCCTGGGCGCAGGACAACTACTTGATGCTCAACCCCAACAG	596	
Db	417	GCGCGCGACGACAAG	AAGCC---GGGCAAGGAGCTTCTGACGCTGACCCGCGACTC	473	
Qy	597	CGAGTACACCTTCG	CCGACCGGGTCTTTCCGCGCGCGCGCGCAAGCGCTCAGCCACCGCGC	656	
Db	474	CTACAA	CATGTTCGAAGAACG	CGAGCTTCTCGCGCGCGCGCGCGCTTCAAGAAGAGGA	533
Qy	657	GCGCGTCCCCGCGCC	CGGGGTGCGGCCGCGAGAGAGGCCCC-----GGCGCTCCCGCGCGCC	711	
Db	534	CGGCTTCAAGGACAAG	GAGGAGAAGGACAGCGCTGCACTCAAGAGAGCGCGCCCGCGCCCGG	593	

```

RESULT 10
US-10-007-280A-115
; Sequence 115, Application US/10007280A
; Publication No. US20030059784A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Chenghua, Liu
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0257
; CURRENT APPLICATION NUMBER: US/10/007,280A
; CURRENT FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,640
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 115
; LENGTH: 2753
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-007-280A-115

```

Query Match	13.3%;	Score 161.4;	DB 14;	Length 2753;
Best Local Similarity	60.2%;	Pred. No. 4.6e-25;		
Matches 305;	Conservative 0;	Mismatches 196;	Indels 6;	Gaps 2;
Qy	207	CGGCGCGGCGCGAGGAGCGATCCCGGAGCAGCTGCTGCACGGTGTGGTGGCGGAGGG	266	
Db	256	CGAGCCCGAGGACCGCGCGGGGCGCTGTGGCCCGAGAGNCCGGTCGCACAGTCAGGA	315	
Qy	267	CGCGGAGCGCCGGGGCGCGCGGGCGCAGCGCGCGGGGAGCGCGGAGGGTGTGCAG	326	
Db	316	GCCAGAGGGCGCGCGCGAGCCCGAGCAAGGGCGTGGGGGTGGCGCGGAGAGCCCC	375	
Qy	327	CAGCAAGCCATATACGGCGGCGGCCAAGCCCCCTACTCGTACATCGCGCTCATCGCCAT	386	
Db	376	GGAGAGCCCGGACCCGGCGAG---AAGCCCCGTACTCGTACGTGGGGCTCATCGCAT	432	
Qy	387	GGCCATCCGCGAATCGCGCGGGCGGGCGCTTGACGCTGGCGAGATCAACGAGTACTCAT	446	
Db	433	GGGGATCCGGAGAGCGCGGAGAGAGGCTCACGCTGTCCGGCATATCCAGTACATCAT	492	

QY 519 CGAGTCTTCTGCAAGGTCCTGCGACCCCTCGCGGCCCTGGGCAAGGCAACTACTG 578
DB 381 CGAGTCTTCTGCAAGGTCCTGCGACCCCTGCGACCCCTGGGCAAGGCAACTACTG 437
QY 579 GATGTCTAACCCCAAGAGGAGTACACCTTTCGCGGAGCGGGTCTTTCGCGCGCGCGCAA 638
DB 438 GACCTCTGAGACCGGAGTCTTACAAATGTTTCGAGAACGCGAGCTTCTTCTGCGCGCGCGG 497
QY 639 GCGCTCTACGACCCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 698
DB 498 GCGCTTCAAAAGAGAGGAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 557
QY 699 CTTCCCGG 758
DB 558 GCGG 617
QY 759 CCGCGCGAGGAGGCGG 818
DB 618 GCGCGAGAGAGAGGAGGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 677
QY 819 CAGCATCTTTCGCAAGGCGCTTTCGCGAGCGCTGCGCTCAGGAGACAGCGCGCGCGCGG 878
DB 678 CAGGTGGAGAGCGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 737
QY 879 GCTTCAAGTGGGCGG 938
DB 738 CAGCGG 797
QY 939 GCGCGCGCTGCAAGGCGCTTCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGG 995
DB 798 GCTGCTGCTTCAAGGTGGAGAAATATATGACCTTTCGAGAGCTGCGCGCGCGGAG 857
QY 996 GCTGGGCGCGCGAGGCGGAGGTCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCT 1055
DB 858 GCTGAGCGG 917
QY 1056 CCGG 1076
DB 918 CCGG 938

RESULT 13

US-09-963-285-8
; Sequence 8, Application US/09963285
; Patent No. US2002090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbeck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3289
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-963-285-9

Query Match 12.7%; Score 153.8; DB 9; Length 3289;
Best Local Similarity 52.4%; Pred. No. 1.7e-23;
Matches 388; Conservative 0; Mismatches 347; Indels 6; Gaps 2;

QY 339 TAGCGG 398
DB 1397 TAGGACCTGGTGAAGCGCGCGCTACAGCTACATCGCGCTCATCACCATGCCATCCAGAA 1456
QY 399 CTGCGCGCGCGCGCGCGCTTGAAGCTGCGCGAGATCAAGAGTACTCTATGCGGCAAGTTCCC 458
DB 1457 CGCGCGCGGAGAAGATCACTTGAAGGCACTACCGAGTTTATCATGACCGCTTCCC 1516
QY 459 CTTTTTCCGCGCGAGCTTACACGCGCTGGCGCAACTCCGTGCGCGCAACAACCTTTTCGCTCAA 518
DB 1517 CTTTACCGGAGAGCAAGCAGAGGCTGGCAGACAGCATCCGCGCAACCTCTCGCTCAA 1576
QY 519 CGACTGTTCTGTCAGAGTGTGTCGCGACCCCTCGCGCGCTTGGGCGAGAGCAACTACTG 578
DB 1577 CGAGTGTCTTCTCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1633
QY 579 GATGCTCAACCCCAAGAGGAGTACCTTTCGCGCGCGCGGCTTCTCGCGCGCGCGCGCAA 638
DB 1634 GACCTGGACCGGACTCTTCAACATGTTTCGAGAGCGGAGTCTCTCGCGCGCGCGG 1693
QY 639 GCGCTCAGCCACCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 698
DB 1694 GCGCTTCAAAAGAGAGAGCGTGTCAAGGAGAGAGAGCGCGCGCGCGCGCGG 1753
QY 699 CTTCCCGG 758
DB 1754 GCGG 1813
QY 759 CGCGCGCGAGGAGCGG 818
DB 1814 GCGCGAGAGAGGTGTGTATCAAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGG 1873
QY 819 CAGCATCTTTCGCAAGGCGCTTTCGCGAGCGCTGCGCTCAGGAGACAGCGCGCGCGCGG 878
DB 1874 CAAAGTGGAGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1933
QY 879 GCTTCAAGTGGGCGG 938
DB 1934 CACGCGCGCGCGCTCCCGCGAGCGTTTCTGCTGCGCGAGACACACCGCGCGCGCGCGG 1993
QY 939 GCGCGCGCTGAGGCGCGCTGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGG 995
DB 1994 GCTGCTGCTTTCAGCGTGGAGAACATCATGACCTTTCGGAACCTGCGCGCGCGCGGAG 2053
QY 996 GCTGGGCGCGCGAGCGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1055
DB 2054 GCTGAGCGG 2113
QY 1056 CCGG 1076
DB 2114 CCGG 2134

RESULT 14

US-09-963-285-1
; Sequence 1, Application US/09963285
; Patent No. US2002090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbeck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2235) ... (3737)
US-09-963-285-1

Query Match      12.7%; Score 153.8; DB 9; Length 6458;
Best Local Similarity 52.4%; Pred. No. 1.6e-23;
Matches 388; Conservative 0; Mismatches 347; Indels 6; Gaps 2;

Qy 339 TACGCGCGCGCCCAAGCCCTTACTGCTATCATCGGCTCATCGCCATGCGCATCGCGCA 398
Db      |||
Qy 2435 TAAGGACCTGGTGAAGCGCCCTACAGTACATCGGCTCATCACCATGGCCATCCAGAA 2494
Db      |||
Qy 399 CTGCGCGCGCGCGGCTTGAAGCTGGCGGAGATCAAGAGTACTCATGGGCAAGTTCC 458
Db      |||
Qy 2495 CGCGCCCGAGAAGAAGATCACTTGAACGGCATCTACCAATTCATGAGACGGCTTCC 2554
Db      |||
Qy 459 CTTTTTCGCGCGGAGCTACACGGGCTGGCGCAACTCGTGGCGCACAACTTTTCGCTCAA 518
Db      |||
Qy 2555 CTTCACCGGAGAGACAAGCAGGGCTGGCAGACAGCATCCGCCACAACTTCGCTCAA 2614
Db      |||
Qy 519 CGACTGCTTGTCAAGTGTGTGCGGACCCCTCGCGGCCCTGGGGCCCAAGCAACTACTG 578
Db      |||
Qy 2615 CGAGTGTCTTGTCAAGTGTGCGCGGACGACCAAGAGCC--GGCAAGGGCGAGTTACTG 2671
Db      |||
Qy 579 GATGCTCAACCCCAACAGCAGTACACTTTCGCGGAGGGGTCTTCGCGCGCGCGCA 638
Db      |||
Qy 2672 GACCTTGAACCGGACTCTTCAACATGTTGAGAGAGCGAGTTTCTGGGGCGCGGG 2731
Db      |||
Qy 639 GCGCTCAGCAACCGCGCGCGGTTCGCGCGCGCGGCTGCGGCCGAGAGGCGCGCGG 698
Db      |||
Qy 2732 GCGCTTCAAAAGAAAGACGTGTCAGGAGAGAGAGAGCGGGCCACCTCAAGGAGCC 2791
Db      |||
Qy 699 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 758
Db      |||
Qy 2792 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2851
Db      |||
Qy 759 CGCGCGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818
Db      |||
Qy 2852 GCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2911
Db      |||
Qy 819 CAGCATCTCGGCAAGCCCTTCGCGAGCGCTCGCTCAGGAGACAGCGCGCGCGCGCG 878
Db      |||
Qy 2912 CAAGGTGAGACGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2971
Db      |||
Qy 879 GCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938
Db      |||
Qy 2972 CACGCGCGCGCGCTCCCGCGAGCGTTGCTGCGCGGAGACACAGCGCGCGCGCGCG 3031
Db      |||
Qy 939 GCGCGCGCTGAGGCGCGCTGCGGCTTCTGCGGCTAGCGCGCGCGCGCGCGCGCGCG 995
Db      |||
Qy 3032 GCTGCTGCTGCTTCAAGCGTGAGAGAACATCATGACCTTCGCAACGTCGCCCGCGCGGAG 3091
Db      |||
Qy 996 GCTGGGCGCGCGGAGCGCGAGGTGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1055
Db      |||
Qy 3092 GCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
Db      |||
Qy 1056 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1076
Db      |||
Qy 3152 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3172
Db      |||
```

RESULT 15

```
US-10-027-632-138703
; Sequence 138703, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138703
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138703
```

```
Query Match      12.1%; Score 145.8; DB 15; Length 425;
Best Local Similarity 65.9%; Pred. No. 1.1e-21;
Matches 226; Conservative 1; Mismatches 113; Indels 3; Gaps 1;

Qy 342 GCGGCGCGCCCAAGCCCTTACTGCTATCGGCTCATCGCCATGGCCATCGCGGACTC 401
Db      |||
Qy 81  GARGCGGCAAGCCCTTACTGCTATCGGCTCATCGCCATGGCCATCGCGGACTC 401
Db      |||
Qy 402 GCGGCGCGCGCGCTTGAAGCTGGCGGAGATCAAGAGTACCTCATGGGCAAGTTCCGCTT 461
Db      |||
Qy 141 CCGGCAACAAGCGCTCAGCTCAGCGCATCTGGCGCTTCAATAGTGGCGCTTCCGCTA 200
Db      |||
Qy 462 TTTCCGCGGAGCTACAGGCGCTGGCGCACTCCGTCGCGCACACCTTTCGCTCAACGA 521
Db      |||
Qy 201 CTACCGCGCGAAGTTCCCGCGCTGGCAGAACAGCATCCGCCAACCTCTCGCTGAACGA 260
Db      |||
Qy 522 CTGCTTCTCAAGTGTGCTGCGGACCCCTCGCGGCGCTGGGGCGAAGGACAACTACTGAT 581
Db      |||
Qy 261 CTGCTTCTCAAGATCCCGCGGAGC--CGGGCCACCCAGGCAAGGCACTACTGAG 317
Db      |||
Qy 582 GCTCAACCCCAAGAGGATACACTTTCGCGGAGCGGCTTTCGCGCGCGCGCGCGCG 641
Db      |||
Qy 318 CTTGACCCCGCGCTCCCGAGGACATGTTTCGACATGGGAGCTTCTCGCGCGGTAGGAAGCG 377
Db      |||
Qy 642 CTTGACCCCGCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 684
Db      |||
Qy 378 TTTCAAGCGCCACCAACTGACCCCGGGAGGCCACCTGCGCCAC 420
Db      |||
```

Search completed: March 31, 2004, 02:26:07
Job time : 450 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 22:45:27 ; Search time 2984 Seconds

(without alignments)
12098.987 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggaggtgttgt.....tggagacgtctctgcttga 1209

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estl:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gas_hum:*

18: em_gas_inv:*

19: em_gas_pln:*

20: em_gas_vrt:*

21: em_gas_fun:*

22: em_gas_mam:*

23: em_gas_mus:*

24: em_gas_pro:*

25: em_gas_rod:*

26: em_gas_pig:*

27: em_gas_vrl:*

28: gb_gsel:*

29: gb_gsel2:*

Result No.	Score	Match	Length	DB	ID	Description
1	537.6	44.5	904	13	BQ922461	BQ922461 AGENCOURT
C 2	350.8	29.0	559	9	AI169632	AI169632 EST15519
C 3	309	25.6	514	10	BF282916	BF282916 EST447507
4	301.2	24.9	398	9	AI586081	AI586081 v092c05.x

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```

C 5 269.2 22.3 425 10 BF290883
6 241.2 20.0 982 13 BQ30047
7 216.8 17.9 682 9 AL636071
C 8 215.2 17.8 518 10 BB637563
9 211.2 17.5 232 12 B1402320
10 209.8 17.4 564 13 BQ449254
11 209.8 17.4 946 14 CD754754
12 209.8 17.4 1117 14 CK028187
13 208.2 17.2 799 14 CF997203
14 205 17.0 881 13 BX084216
15 202.2 16.7 581 13 BQ285521
16 196.8 16.3 804 14 CK030218
C 17 183.2 15.2 688 12 B1443539
18 180.4 14.9 878 29 CN3044M2
19 179.2 14.8 246 29 CE688385
20 174.2 14.4 646 14 CB576131
21 170.8 14.1 542 14 CD282719
22 166.8 13.8 406 12 BM510453
23 163 13.5 885 11 BC019896
24 158 13.1 498 13 BX282141
25 158 13.1 635 13 BY125661
26 154.2 12.8 626 13 BY13753
27 153.4 12.7 842 12 B1254526
28 153.4 12.7 2293 11 BC055774
C 29 152.2 12.6 430 10 AW532942
C 30 152 12.6 445 10 AW236501
31 151.6 12.5 585 13 BQ480599
32 151.2 12.5 556 13 BQ189385
33 150.8 12.5 471 10 BE501220
34 150.4 12.4 2443 11 BC035020
35 149.6 12.4 356 12 BM529336
36 149.6 12.4 458 12 BM529696
37 149.6 12.4 471 12 BM529436
38 149.6 12.4 476 12 BM574346
39 149.6 12.4 483 12 BM529717
40 149.6 12.4 498 12 BM529729
41 149.6 12.4 541 12 BM572533
42 149.6 12.4 554 12 BM574905
43 149.6 12.4 556 12 BM573325
44 149.6 12.4 557 12 BM531626
45 149.6 12.4 557 12 BM532671

```

ALIGNMENTS

```

RESULT 1
BQ922461
LOCUS      BQ922461      904 bp      mRNA      linear      EST 20-AUG-2002
DEFINITION AGENCOURT 8963457 Lupski sciatic_nerve Homo sapiens cDNA clone
IMAGE:6200329.5', mRNA sequence.
ACCESSION  BQ922461.1 GI:22337492
VERSION     BQ922461.1
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 904)
            NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-rc@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Agencourt Bioscience Corporation.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMA13615 row: a column: 02
            High quality sequence stop: 523.

```

```
FEATURES
Source
Location/Qualifiers
1..904
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6200329"
/sex="male"
/tissue_type="sciatic nerve"
/dev_stage="adult, 70 yr"
/lab_hosts="DRI08"
/clone_lib="Lupski_sciatic nerve"
/notes="vector: pCMV-SPORT6 (Life Technologies); Site 1:
NotI; Site 2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGCCG-3' and
5'-GACTAGTCTTAGTCGCGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.87 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
```

```

Query Match 44.5%; Score 537.6; DB 13; Length 904;
Best Local Similarity 87.3%; Pred. No. 6.9e-52;
Matches 610; Conservative 0; Mismatches 87; Indels 2; Gaps 2;
QY 488 GCAACTCGGTGGCGCAACACTTTCGCTCAACGACTGCTTCGTCGAAGTGCTCGCGGACC 547
Db 1 GCAACTCGGTGGCGCAACACTTTCGCTCAACGACTGCTTCGTCGAAGTGCTCGCGGACC 60
QY 548 CCTCGCGGCGCTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGGAGTACACCT 607
Db 61 CCTCGCGGCGCTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGGAGTACACCT 120
QY 608 TCGCGGAGGGGCTCTTCGCGCGCGCGCGCAAGGCGCTCAGCCACCGCGCGCGTCCCG 667
Db 121 TCGCGGAGGGGCTCTTCGCGCGCGCGCGCAAGGCGCTCAGCCACCGCGCGCGTCCCG 190
QY 668 CGCCCGGCGCTGCGCGGAGAGCGCCCGGCGCTTCCCGCGCGCGCGCGCGCGCGCG 727
Db 181 CGCCCGGCGCTGCGCGGAGAGCGCCCGGCGCTTCCCGCGCGCGCGCGCGCGCGCG 240
QY 728 CGCGCGCGGCTCGCGCGGATGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787
Db 241 CGCGCGCGGCTCGCGCGGATGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 788 CGGCAAGATTCTCCAGGTCTCTTCGCCATCGACAGCATCTTCGCGCAAGCCCTTCGCGAGCC 847
Db 301 CGGCAAGATTCTCCAGGTCTCTTCGCCATCGACAGCATCTTCGCGCAAGCCCTTCGCGAGCC 360
QY 848 GTGCGCTCAGGACACGGCGCGCGGAGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCGCG 907
Db 361 GCGCGCTCAGGACACGGCGCGCGGAGAGCGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCG 420
QY 908 GCGTGCCTGCTTCCCGCGCTCTCTCCCGCGCGCGCGCTTCAGGGCGCGCGCGCGCGCTCT 967
Db 421 GCGTGCCTGCTTCCCGCGCTCTCTCCCGCGCGCGCGCTTCAGGGCGCGCGCGCGCGCTCT 480
QY 968 GCGGCTACGGCGCGGCGAGCGCGCGGCTTGGGCGCGCGCGGAGCGCGAGGTGCCACCGA 1027
Db 481 GCGGCTACGGCGCGGCGAGCGCGCGGCTTGGGCGCGCGGAGCGCGAGGTGCCACCGG 540
QY 1028 CGCGCGCGCGCGCTCTGCTTGCATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
Db 541 TC-CCCGGCGCGCGCGCTTGCATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 598
QY 1088 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147
Db 599 CCGATCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658
QY 1148 CGGCGCTAGTCCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
```

```

Db 659 TTGTTCCCGCCCTCCCGCTATCCCGCGCTTTGTTC 697

RESULT 2
A1169632/c
LOCUS
DEFINITION
EST215519 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
RKBV46 3' end, mRNA sequence.
ACCESSION
A1169632
VERSION
A1169632.1 GI:3709672
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
REFERENCE
1 (bases 1 to 559)
Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL
Unpublished (1998)
COMMENT
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
1..559
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2027517"
/db_xref="taxon:10118"
/clone="RKBV46"
/clone_lib="Normalized rat kidney, Bento Soares"
/notes="Organ: kidney; Vector: pT73Pac; Site 1: EcoRI;
Site 2: NotI"
ORIGIN
Query Match 29.0%; Score 350.8; DB 9; Length 559;
Best Local Similarity 80.2%; Pred. No. 1e-30;
Matches 451; Conservative 0; Mismatches 102; Indels 9; Gaps 3;
QY 590 CCAACAGAGGTACACCTTCGCGGCGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 649
Db 559 CCAACAGAGGTACACCTTCGCGGCGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
QY 650 ACCGCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 709
Db 499 ACCGCGCGCGGTTCGCGATCGCGGCTACGCGCGGAGAGCGCGCGCGCGCGCGCGCG 440
QY 710 ----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 766
Db 439 GGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
QY 767 AGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
Db 379 AGGAGGAGCGGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
QY 827 TCGCAAGCGCTTCGCGAGCGGTTCGCGTTCAGGACACCGCGCGCGCGCGCGCGCGCGCG 886
Db 319 TCAGCAAGCGCTTCGCGAGCGGTTCGCGTTCAGGACACCGCGCGCGCGCGCGCGCGCG 260
QY 887 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
Db 259 GGAGCGCTGCTCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
QY 947 GCAGGCGCGCTGCTCGCGTTCGCGTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006
Db 199 GCGGTGCGCTGCTCGCGTTCGCGTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 140
```

QY 1007 GCGAGCGAGGTGCGACCGCGCGCGCCCTCTCTGCTTGACCTCTCCCGGGGGG 1066
 Db 139 GCGGGCGCGAGGTGCAAC---CGGGCGCGCCCTGTGCTGGCGCCCTCTCCACCGGG 83
 QY 1067 CCCCCGCCCAAGCACTCCGAGGCCGGG---CGGCGGGCGGGCGCGCACTGTACTGCCCC 1123
 Db 82 CCCCAGCCAAAGCCATTTCGAGTCCGGAGACCGCGGGCGCGCACTGTACTGCCCC 23
 QY 1124 TCGGCTGCCCGCGAGCCCTGCA 1145
 Db 22 TACGGCTGCCCGAGCCCTGCA 1

RESULT 3
 LOCUS BF282916/c
 DEFINITION BF282916 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
 ACCESSION BF282916
 VERSION BF282916.1 GI:11213986
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 514)
 Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
 Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
 Generation of ESTs from Normalized Rat Embryo, Bento Soares
 Unpublished (2000)
 Other ESTs: EST348098
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igr.org
 This clone is available through the ATCC, contact the ATCC
 tel#703-365-2700 for further information.

FEATURES
 Location/Qualifiers
 1..514
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="RGIDU92"
 /tissue_type="mixed tissue"
 /lab_host="DH5-alpha"
 /clone_lib="Rat Gene Index, normalized rat, Rattus
 norvegicus cDNA"
 /note="Vector: pT37Pac; Site 1: EcoRI; Site 2: NotI;
 Combination of ROV, RBR, RKL, RLI, RPL, RLU, REM, RMU,
 RSP, RHE, RPC, RPN"

ORIGIN
 Query Match 25.6%; Score 309; DB 10; Length 514;
 Best Local Similarity 78.9%; Pred. No. 5.4e-26;
 Matches 408; Conservative 0; Mismatches 100; Indels 9; Gaps 3;
 QY 636 CAAGCGCTCAGCCACCGCGCGCGTCCCGCGCGCGGTGCGCGCGAGAGGCC 695
 Db 514 CAAGCGCTCAGCCACCGCGCGCGTCCCGCATCGGGGTACGCGCGAGAGGCC 455
 QY 696 GGGCTCTCCCGCG---CCCGCGCGCGCGCGCGCGCGCGCGCGCGCATGCG 752
 Db 454 ACCCGACTCGGGGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 395
 QY 753 CTGCGCGCGCGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 812
 Db 394 CTGCGCGCGCGCGCGAGGAGGCGTCCAGCGCGCGCGCGCGCGCGCGCG 335
 QY 813 CATCGACATCTCTGCGAGCGCTTCGCGAGCGGTGCGCTCAGGACAGCGCGCG 872

Db 334 CATCGACAGATCCTCAGCAAGCGGTTTCGACGCGCGCGCGCGCGCGCTCTGGG 275
 QY 873 GAGACGCTTCACTGGGGCGCGCGCGCTGCGCGCGCTGCGCGCGTTCCTTCTCT 932
 Db 274 GGTGCACTACCTTGAGCGCTGCTCCCTGCGCGCGCGCTGCGCGCGCTATCC 215
 QY 933 CCGCGCGCGCGCGCTGAGGGCGCTGCTGCGCGCTCTGCGCGTACGGCGCGCGCG 992
 Db 214 TCCGCGCTGCTGCGCGGTGCTGCTGCGCGCTCTGCTTACGGCGCGCGCG 155
 QY 993 CGCGCTGGCGCGCGCGAGCGCGAGGTGCGACCGAGCGCGCGCGCTTCTCTGTG 1052
 Db 154 GCTGCTGGCGTCCGCGCGCGCGCGAGGTGCAAC---CGCGCGCGCGCGCTGT 98
 QY 1053 TCTCCCG 1109
 Db 97 CCTTCACCG 38
 QY 1110 CTTGACTGCG 1146
 Db 37 CTTGACTGCG 1

AI586081 398 bp mRNA linear EST 06-APR-1999
 v192c05.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1209608 3' similar to gb:U36760 Mus musculus brain factor-1
 (MOUSE); mRNA sequence.
 ACCESSION AI586081
 VERSION AI586081.1 GI:4571978
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 398)
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Persson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
 Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.lml.gov) for further information.
 MGI:645952
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 387.
 Location/Qualifiers
 1..398
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1209608"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="Stratagene mouse skin (#937313)"
 /notes="Organ: skin; Vector: pBluescript SK-; Site 1:
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

/db_xref="taxon:9606"
 /clones="IMAGE:6010412"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 68"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 20.0%; Score 241.2; DB 13; Length 982;
 Best Local Similarity 95.7%; Pred. No. 2.9e-18;
 Matches 270; Conservative 0; Mismatches 8; Indels 4; Gaps 2;
 QY 1 ATGAAGTTGGAGGTTTCTGCTCCCTCGCGCGGCCACAGCGGAGGAGTGCCTG 60
 Db 265 ATGAAGTTGGAGGTTTCTGCTCCCTCGCGCGGCCACAGCGGAGGAGTGCCTG 324
 QY 61 GAGGCGCGGGCGGAGCGAGCGCGCTCCCGCTGTCGCGCGGGGAGAGACTCCCTG 120
 Db 325 GAGGCGCGGGCGGAGCGAGCGCGCTCCCGCTGTCGCGCGGGGAGAGACTCCCTG 384
 QY 121 GCTCAGATGGGGAATGCGCGGCCAA---GCCGTCCGCGGCGCGCGGCGGAGATCG 177
 Db 385 GCTCAGATGGGGAATGCGCGGCCAAACAGCGCGCGCGCGCGGCGGAGATCG 444
 QY 178 CAGGCGCGGCGGAGCGAGTCCGCGGCGCGCGCGCGCGGCGGAGGCGATCCCGCA 237
 Db 445 CCGGCGGCGGAGCGAGTCCGCGGCGCGCGCGCGCGCGGCGGAGGCGATCCCGCA 504
 QY 238 GCAGCTG-CTGCAGCGGTGGTGGCGGAGGCGCGGCGGCGGCGGCGGCGGCGG 278
 Db 505 GCAGCTGCTGCAGCGGTGGTGGCGGAGGCGCGGCGGCGGCGGCGGCGGCGG 546

RESULT 7

AL636071 682 bp mRNA linear EST 07-NOV-2003
 LOCUS AL636071 XGC-neurula Silurana tropicalis cDNA clone Tneu015112 5',
 DEFINITION mRNA sequence.

ACCESSION AL636071.2 GI:38214106

VERSION EST.

SOURCE Silurana tropicalis (western clawed frog)

ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 Xenopodinae; Silurana.

REFERENCE 1 (bases 1 to 682)

Authors Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

Title Sanger Xenopus tropicalis EST project 2001 (11_2003)

Journal Unpublished (2003)

Comment On Nov 7, 2001 this sequence version replaced gi:16788050.

Contact: Huckle E

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: tropesanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dT primed from 5ug of poly A+ RNA from neurula.

ECORI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the

5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli DH10B

Sanger Xenopus tropicalis EST project 2001

TROPICALIS SEQUENCE ID: Tneu015112.p1kSP6

Sequencing primer: SP6.

Location/Qualifiers

1...682

/organism="Silurana tropicalis"

/mol_type="mRNA"

/db_xref="taxon:8364"

FEATURES

source

/clone="Tneu015112"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /clone_lib="XGC-neurula"
 /notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dT primed from 5ug of poly A+ RNA from neurula.
 EORI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 17.9%; Score 216.8; DB 9; Length 682;
 Best Local Similarity 68.0%; Pred. No. 1.6e-15;
 Matches 302; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
 QY 205 GGCGGCGCGGCGGAGGAGGCGATCCCGGAGCGAGCTGCGCAGCGGTGGTGGGGAG 264
 Db 19 GTGAGCTTGGGGGGATGAAGAAGATGAAGTGAAGAAGAGAGGAGGTAAACCCAGAG 78
 QY 265 GGCGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 324
 Db 79 AGAATGGGGTCTCGGCAGATGGTTCTACGCAATCCGAGCACAAATTTGAGGGGGG 138
 QY 325 CGCAGCAGCATATACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 384
 Db 139 AAAACAAAGACTTACACCGCGCTTTTCGAAACCTCTTACTCTTACTCTTACTCTT 198
 QY 385 ATGGCCATCGGACTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
 Db 199 ATGGCTATTAAGACTCAGCCAGTGGCGGCTGACCTTGCAGAGATCAATGACTATCTG 258
 QY 445 ATGGGCAAGTTCCTTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 504
 Db 259 ATGAAGAAGTTCCTTCTTTCAGGGGTAGCTACACGGGCTGGAGGAACTCGGTGCGCAC 318
 QY 505 AACCTTTTCGCTCAACGACTGCTTCGCAAGGTCTCGCGGAGCCCTCGCGGCGGCGGCGGCGG 564
 Db 319 AATCTTTTCCTCAACGACTGCTTCGCAAGGTCTCGCGGAGCCCATCACGGCTTGGGGC 378
 QY 565 AAGCACAATCTGAGTGTCAACCCCAACAGGAGTACACCTTCCCGCAGCGGTCTTC 624
 Db 379 AAGACAACTACTGAGTGTGAACCCCAATAGTAGTACACCTTTCGAGACGCGGCTTC 438
 QY 625 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 648
 Db 439 AGACGAGGAAGGAGCGACTTAAAC 462

RESULT 8

BB637563/c

LOCUS BB637563

DEFINITION BB637563 RIKEN full-length enriched, adult male aorta and vein Mus

musculus cDNA clone A53084C06 5', mRNA sequence.

ACCESSION BB637563

VERSION BB637563.1

KEYWORDS GI:16473338

SOURCE EST.

Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 518)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,

Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,

Kanno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,

Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,

Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayaishizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoreda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,
and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., and
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and
Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues. Location/Qualifiers

FEATURES

source

1. 518
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A530084C06"
/sex="male"
/issue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male aorta
and vein"

/notes="Site 1: Salt; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAAGATCCAGAGCTCTTTTCTTTTCTTTT 3']. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5', GAGAGAGATCTCGATTAAATAATATCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
pUC 1."

ORIGIN

Query Match 17.8%; Score 215.2; DB 10; Length 518;
Best Local Similarity 74.5%; Pred. No. 2.4e-15;
Matches 319; Conservative 0; Mismatches 88; Indels 21; Gaps 3;

QY 1 ATCAATGGAGTGTCTCTCGCGGGCCACGGGGAAGCAGGCGGAGTACCTG 60
DB 413 ATGAATGGAGTTTCTCTCCACCGGAGCCCGGGGAAAATGGGAGGATTG 354
QY 61 GAGGCGCGCGGCGGACGACGCGCTCCCGCTCGCGCGGAGGACACCTCCCTG 120
DB 353 GAGGCGCGGCGGACGACGCGTGCATCTCCACTGTCTCGCGCTGGTGCAGCTCTTA 294

QY 121 GGCTCAGATGGGAGTCTCGCGGCCAA---CCCGTCCCGCGCGCGCGCCAGAGATACG 177
DB 293 GGCTCAGACGGGAGTCTGTGAGCCAAACAGCCCGCGCGCGCGCGGGGATTG 234
QY 178 CA---GGGCGCAGCGGGAACAGAGTGCAGGAGGGGCGCGGAGGAGGCGATCCCG 234
DB 233 GAAGTGGCGGCGCGAGAGAGAAATTCAGTGGCGGCGCGAGCGCCCAAGACGGTCCGGAG 174
QY 235 GCACGACTGCTGCAGCGGTGTGTGGGAGGAGGCGCGGAGGCGCGGCGGCGGCGGCGGCGG 294
DB 173 GCAACTGAT-----GACAGCAGAAACGAGGCGCTCCCGCGGCGAGGCGCGTGC 129
QY 295 GCGGCGCGCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 354
DB 128 GCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 69
QY 355 CCCCCCTACTGCTACATCGGCGCTCATCGCCATGGCCATCCGCGACTCGGCGGCGGCGGCGG 414
DB 68 CCCCCATCTCTCTACATCGCTCTTATCGCATGCGCATCGCGACTCCCGCGGCGGCGGCGG 9
QY 415 TTGACGCT 422
DB 8 CTGACACT 1

RESULT 9

BI402320

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sus scrofa

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 232)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

889548

COMMENT

Contact: Tugle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildree Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktugle@iastate.edu

Oligo-dT track not found, Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,

University of Iowa Clone distribution: clones will be available

through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1. 232

/organism="Sus scrofa"

/mol_type="mRNA"

/db_xref="taxon:9823"

/clone="MI-P-CP0-nwd-f-07-0-UI"

/lab_hosts="DH10B (Life Technologies)"

/clone_lib="MI-P-CP0"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: EcoRI; The MI-P-CP0

library is derived from uterus. For a detailed description

of the library from which this clone was derived, please

visit our web site at <http://pigest.genome.iastate.edu/>.

The procedure used to create this library has been

previously described (Bonaldo, Lennon and Soares, Genome

```

/clone lib="zebrafish fin dev3" regeneration"
"vector" PBK-CMV; Site1: EcoR; Site2: XhoI; 1st
strand cDNA (1.1 kb) was amplified with (GATC)AATGCTCTGAG(T)18 followed
by second strand synthesis and ligated to 5' adapter
(5')-aattcgccagcag-3', 3'-gccggcgctc-5'. cDNA was cloned
directionally (EcoRI/XhoI) into Stratagene zap expressed
lambda phage arms. Maes invivo excision done to obtain
insets in PBK-CMV phagemid."

```

notes" vector: pbk-cmv; Site_1: EcoRI; Site_2: XhoI; 1st strand cDNA primed with (GA)10ACTAGTCTCGAG(T)₁₈, followed by second strand synthesis, and ligated to 5' adapter (5')-aattcgccagag-3', 3'-gccggtctc-5'. cDNA was cloned directionally [EcoRI/XhoI] into Stratagene Zap express lambda phage arms. Mass invivo excision done to obtain inserts in pbk-cmv phagemid."

ORIGIN	Query Match	17.4%;	Score 209.8;	DB 13;	Length 564;
	Best Local Similarity	70.5%;	Pred. No. 9.8e-15;		
	Matches 280;	Conservative	0;	Mismatches 117;	Indels 0;
				Gaps	0;
QY	251	CGGTGTTGGCGAGGCGCGGAGGCGCGGGCGCGGGCCAGCGCGGGCGCGCGCGGGGA	310		
Db	94	CGGTGTCCGACAGGAGGAGCTGGGCTCGGATCGAGACTCGCTGTGGGCGCAGTCCGGCAC	153		
QY	311	GGGGCGAGGGTGCACGCACACAGCCATATACGGCGGGCGCCCAAGCCCCCTACTCTGTTACA	370		
Db	154	CTGTGCCCGCAGACCAAGAGCAACCCCTACCTCGGAGACCCAAACCTCCATATCTTTTACA	213		
QY	371	TGCGCGCTCATCGCCATGCGGCATCCGCGGACTCGCGCGGGCGGGCGCTTTGACGCGCTGGCGGAGA	430		

214	1	CGACCTTATGCCATGGCCATCCCGAGATCCACATCCCGCCGACCTCATCTCTACCGCGGAA	275
431	QY	TCAAACGAGTACTCATGGCGCAAGTCCCTCTTTTCCGCGCGACGCTACACGGGTCTGGCGCA	490
274	Db	TCAACGACTACTCATGAGAGAGTCCCTGTTTTTAGAGGCAGCTACACGGGTCTGGCGCA	333
491	QY	ACTCGGTGGCGCAACCTTTTGGCTCAAGACGCTGCTTCGTCAAGGTGCTCGCGGACCCCT	550

QY	491	ATCTGGTGGCCACACACCTTTTGGCTTAAGACATGCTTCGTCAGAGTGCATCGGACACCTCT	390
Db	334	ATCAGTGGCCATATCTGTCTCTTAAACGATGCTTCTCAAGGTCCTTAAGGATGCTT	393
QY	551	CGCGGCCCTGGGGCAAGGACAACCTACTGGATGCTCAACCCCAACAGGAGTACACCTTCG	610

394 CGAGACCGTGGGAAAGGACAAATTACTGGATGCTGAACCGGCACAGCGAGTACACCTTCG 453

Db

611 CGACGGGGTCTTCGCGCCGCCCGCAAGCGCTCAG 547

QY

454 CGGACGGAGTGTTTCGACAAAGGAAAGCGCATTAG 490

Db

RESULT	11
CD754754	
LOCUS	
DEFINITION	
CD754754	946 bp mRNA linear EST 30-JUN-2003
AGENCORDBT	14618637 NCBI CCAB ZEMB2 Danio rerio cDNA clone

DEFINITION
IMAGE:5964741 5', mRNA sequence.
AGENCOURT 14618637 NCI CGAP ZENB2 Danio rerio cDNA clone
ACCESSION
CD754754
VERSION
CD754754.1 GI:32339041
KEYWORDS
EST.
SOURCE
Danio rerio (zebrafish)

SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 946)

AUTHORS NIH-MGC <http://mgc.mci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Leonard I. Zon, M.D.
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arranged by: The T M A C E Consortium (TINT)

CJNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL4505 row: c column: 12

14000: JEFFERSON TOWN: 5 COLUMBIAN: 12

Db 511 GTTGGAGAAATTCGGTCGCCCAAACTGTGCACTTAACGACTGTTTTCTAAAGTCTCTCC 570
 QY 542 GCGACCCCTCGCGCCCTGGGGCAAGCAGCACTACTGGATGCTCAACCCCAACAGGAGCT 601
 Db 571 GGGACCCCTCGACACCTCGGGAAAGACAACTACTGGATGCTGAATCTTCACAGGAGT 630
 QY 602 ACACCTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCTCA 646
 Db 631 ACACCTTCGCTGACGGAGTGTTCGACGCGAGGAGAAAGCGCATCA 675

RESULT 15
 BQ285521 581 bp mRNA linear EST 14-MAY-2002
 LOCUS fa51a09.y1 zebrafish fin day3 regeneration Danio rerio cDNA clone
 DEFINITION IMAGE5912369 5' similar to SW:HFH1 RAT Q63244 HEPATOCYTE NUCLEAR
 FACTOR 3 FORKHEAD HOMOLOG 1 ; mRNA sequence.

ACCESSION BQ285521
 VERSION BQ285521.1 GI:20654021
 KEYWORDS EST.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio

REFERENCE 1 (bases 1 to 581)
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
 Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
 Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 TITLE WashU Zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu
 cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
 info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 RessourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 487.
 FEATURES
 source 1..581
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:5912369"
 /sex="mixed male and female"
 /tissue_type="3 day fin regenerates"
 /lab_host="E. coli XL0LR"
 /clone_lib="zebrafish fin day3 regeneration"
 /note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; 1st
 strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed
 by second strand synthesis, and ligated to 5' adapter
 (5')-aattggcagcag-3', 3'-gccgtgttc-5'. cDNA was cloned
 directionally (EcoRI/XhoI) into Stratagene Zap express
 lambda phage arms. Mass in vivo excision done to obtain
 inserts in pBK-CMV phagemid."

ORIGIN

Query Match 16.7%; Score 202.2; DB 13; Length 581;
 Best Local Similarity 77.1%; Pred. No. 7.2e-14;
 Matches 246; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 329 GCAAGCCATATACGCGGGGCCCAAGCCCCCTTACTCGTACATCGCGCTCATCGCCATGG 389
 Db 39 GCAAAACCCCTACACTCGGAGACCCAAACCTCCATACTTTACATGGCACTTATCGCCATGG 98
 QY 389 CCATCCGCGACTCGCGGGCGGCTTGACGCTGGCGGAGATCAACGAGTACCTCATGG 448
 Db 99 CCATCCGAGACTCCAACTCCCGCGCGACTCACTTAGCGGAAATCAACGACTACTCATGA 158
 QY 449 GCAAGTTCCTCCCTTTTTCGCGCGCGAGCTACACGGGCTGGCGCAACTCCGTCGCGCACCAACC 508
 Db 159 AGAAGTTCCTCCGTTTTTTAGAGGCGAGTACACCGGCTGGAGGAACTCAGTGGCCATAATC 218
 QY 509 TTTGCTCAACGACTGCTTCTCAAGGTGCTGCGGACCCCTCGGGCCCTGGGGCAAGG 568
 Db 219 TGTCTCTAAACGACTGCTTTCTCAAGGCTTTAAAGGATCCTTCGAGACCGTGGGAAAGG 278
 QY 569 ACAACTACTGGATGCTCAACCCCAACAGCGAGTACACCTTCGCGGAGCGGGGTCTTCGCGCC 628
 Db 279 ACAATTACTGGATGCTGAACCCCGCACAGCGAGTACACCTTCGCGGAGCGAGTGTTCGCA 338
 QY 629 GCGCGCGCAAGCGCTCAG 647
 Db 339 GAAGGAGAAAGCGCATTAG 357

Search completed: March 31, 2004, 01:11:25
 Job time : 2989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 01:13:48 ; Search time 4679 Seconds
(without alignments)
11199.343 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggagtggttgctg.....tggagagcgtcctagcttga 1209

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Length	DB ID	Description
1	1209	100.0	1448	9	AF153341 Homo sapi
2	951	78.7	62761	9	AL499606 Human DNA
3	923	76.3	2322	9	BC053850 Homo sapi
4	923	76.3	4109	9	AF225950 Homo sapi
5	667	55.2	182152	2	AC067929 Homo sapi
6	554	45.8	178168	2	AC012350 Homo sapi
7	406	33.6	74969	2	AC016289 Homo sapi
8	146	12.1	182152	2	AC067929 Homo sapi
9	77	6.4	1513	10	AF154426 Mus muscu
10	77	6.4	2046	10	BC047155 Mus muscu
11	77	6.4	4763	10	AF010405 Mus muscu
12	77	6.4	187709	10	AL589738 Mouse DNA
13	53	4.4	1576	10	AF153193 Rattus ra
14	53	4.4	1760	10	RATHFH1
15	53	4.4	195184	2	AC119457
16	32	2.6	318	9	HSU42990
17	32	2.6	697	9	HSA337367
18	32	2.6	2011	9	AF275722 Homo sapi
19	32	2.6	106239	9	AL607122 Human DNA
20	32	2.6	153604	2	AC022754 Homo sapi
21	30	2.5	1038	5	AF072889 Xenopus l
22	30	2.5	1298	9	AF452723 Homo sapi
23	30	2.5	1817	5	XLAC11652
24	30	2.5	2068	9	AK122950 Homo sapi
25	30	2.5	2164	9	AY344639 Homo sapi
26	30	2.5	2207	9	AY345866 Pan trogl
27	30	2.5	2212	9	AY345859 Pongo pyg
28	30	2.5	2212	9	AY345860 Pongo pyg
29	30	2.5	2219	9	AY344641 Homo sapi
30	30	2.5	2219	9	AY344642 Homo sapi
31	30	2.5	2219	9	AY345862 Gorilla g
32	30	2.5	2219	9	AY345864 Pan trogl
33	30	2.5	2220	9	AY345861 Gorilla g
34	30	2.5	2487	9	AK125398 Homo sapi
35	30	2.5	3107	9	AF343005 Homo sapi
36	30	2.5	136659	9	AL772360 Human DNA
37	30	2.5	148489	2	AL773528 Homo sapi
38	30	2.5	158187	9	AL512605 Human DNA
39	30	2.5	164799	2	AC127362 Homo sapi
40	30	2.5	166636	2	AC140869 Homo sapi
41	30	2.5	172991	9	AL353608 Human DNA
42	30	2.5	174892	2	AL135795 Homo sapi
43	30	2.5	176734	9	HSA395114 Human DNA
44	30	2.5	176942	2	AC141462 Homo sapi
45	30	2.5	179342	9	BX255923 Human DNA

ALIGNMENTS

RESULT 1
AF153341
LOCUS Homo sapiens winged helix/forkhead transcription factor (HFH1)
DEFINITION 1448 bp DNA linear PRI 24-AUG-2001
ACCESSION AF153341
VERSION AF153341.1 GI:8489092
KEYWORDS Gene, complete cds.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1448)
AUTHORS Hong,H.K., Noveroske,J.K., Headon,D.J., Liu,T., Sy,M.S.,
Justice,M.J. and Chakravarti,A.

TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice
 JOURNAL Genesis 29 (4), 163-171 (2001)
 MEDLINE 21207067
 PUBMED 11309849
 REFERENCE 2 (bases 1 to 1448)
 AUTHORS Hong, H.-K., Noveroske, J.K., Justice, M.J. and Chakravarti, A.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAY-1999) Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA
 FEATURES
 source
 1. .1448
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p25"
 <233. .>1441
 /gene="HFH1"
 <233. .>1441
 /gene="HFH1"
 /product="winged helix/forkhead transcription factor"
 233. .1441
 /gene="HFH1"
 /codon_start=1
 /product="winged helix/forkhead transcription factor"
 /protein_id="AA175586.1"
 /db_xref="GI:8489093"
 /translation="MKLEVFVRAAHGDKQSLLEGAGSDAPSLAAGDSDLSGSDG
 DCAAKSAGSGARDTDCGEGAGGPGAEAEIPAAIAAAVVAEGAEAGAGPAGGA
 GSGEGARSPYRPPKPYSLTALIMARDAGGRLTAEINEYLMGKFRPPRGSYT
 GWNSVRHLSLNDKPVKLRDPSRPWGDKNYMLNPNSEYTFADGVFRRRKRLSHR
 APVPAQLPEEPAGCPAAPPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
 RKFPRRLRDTPAGTTLQWGAAPCPPLPAFPALLPAAPCRALLPLCIYAGEPARLG
 AREAVFPPTAPILLAPLPAAPAKPLRPAAGGAHLVCPRLPALQALVRRPGPH
 LSPVETLLA"
 ORIGIN
 Query Match 100.0%; Score 1209; DB 9; Length 1448;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAAGTTGAGGTGTTGCTCCCTCGCGGCGCCACGGGGAACAAGCAGGCGAGTGACCTG 60
 DB 233 ATGAAGTTGAGGTGTTGCTCCCTCGCGGCGCCACGGGGAACAAGCAGGCGAGTGACCTG 292
 QY 61 GAGGGCGGGCGGCGGAGCGAGCGCGCTCCCGCTGTCGCGGCGGAGACGACCTCCCTG 120
 DB 293 GAGGGCGGGCGGCGGAGCGAGCGCGCTCCCGCTGTCGCGGCGGAGACGACCTCCCTG 352
 QY 121 GGCTCAGATGGGACTGCGCGGCGCAAGCCGTCGCGGCGGCGGCGCGCCAGAGATACGCAG 180
 DB 353 GGCTCAGATGGGACTGCGCGGCGCAAGCCGTCGCGGCGGCGGCGCGCCAGAGATACGCAG 412
 QY 181 GGCGACGGCGAACAAGAGTCGGAGAGCGGGCGGGCGGGCGGAGAGCGGATCCGGCAGCA 240
 DB 413 GGCGACGGCGAACAAGAGTCGGAGAGCGGGCGGGCGGGCGGAGAGCGGATCCGGCAGCA 472
 QY 241 GCTGCTGACGCGTGGTGGCGGAGGGCGGAGGCGGGCGGGCGGGCGGCGCGGCGGCG 300
 DB 473 GCTGCTGACGCGTGGTGGCGGAGGGCGGAGGCGGGCGGGCGGGCGGCGCGGCGGCG 532
 QY 301 GGCGCGGGAGCGGCGAGGTGCACGAGCAAGCCATATACGCGGCGGGCGGCGGCGGCGG 360
 DB 533 GGCGCGGGAGCGGCGAGGTGCACGAGCAAGCCATATACGCGGCGGGCGGCGGCGGCGG 592
 QY 361 TACTCGTACATCGCGCTCATCGCATCGGATCGGCGGCTCGGCGGGCGGGCGGCTTGAGC 420
 DB 593 TACTCGTACATCGCGCTCATCGCATCGGATCGGCGGCTCGGCGGGCGGGCGGCTTGAGC 652
 QY 421 CTGGCGGAGATCAACAGTACCTCATGGCAAGTTCCTCTTTTCCGCGGCGAGCTACAGC 480
 DB 653 CTGGCGGAGATCAACAGTACCTCATGGCAAGTTCCTCTTTTCCGCGGCGAGCTACAGC 712

QY 481 GGCTGGCGCAACTCCGTGCGCCCAACACTTTCGCTCAACGACTGCTTCGTCGAAGTGCTG 540
 DB 713 GGCTGGCGCAACTCCGTGCGCCCAACACTTTCGCTCAACGACTGCTTCGTCGAAGTGCTG 772
 QY 541 CGGACCCCTCGCGGCCCTGGGCAAGGACAACACTTCTGGATGCTCAACCCCAAGAGCGAG 600
 DB 773 CGGACCCCTCGCGGCCCTGGGCAAGGACAACACTTCTGGATGCTCAACCCCAAGAGCGAG 832
 QY 601 TACACCTTTCGCGCGGCGGCTTTCGCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCGCG 660
 DB 833 TACACCTTTCGCGCGGCGGCTTTCGCGCGCGCGCGCAAGCGCTTCAGCCACCGCGCGCG 892
 QY 561 GTCCCGCGCGCGGCTTTCGCGCGCGGAGAGCGCGGCTTTCGCGCGCGCGCGCGCGCGCC 720
 DB 893 GTCCCGCGCGCGGCTTTCGCGCGCGGAGAGCGCGGCTTTCGCGCGCGCGCGCGCGCGCC 952
 QY 721 GCGCGCGCGCGCGGCTTTCGCGCGCGCATGCGCTTTCGCGCGCGCGCGCGAGAGAGCGCGCC 780
 DB 953 GCGCGCGCGCGCGGCTTTCGCGCGCGCATGCGCTTTCGCGCGCGCGCGAGAGAGCGCGCC 1012
 QY 781 AGCCCGCGGGCAAGTTCCTCAGCTCTTCGCGCATTCGAGCATCTTCGCGAGCCCTTC 840
 DB 1013 AGCCCGCGGGCAAGTTCCTCAGCTCTTCGCGCATTCGAGCATCTTCGCGAGCCCTTC 1072
 QY 841 CGCAGCGCTCGCTCAGGAGACAGCGGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCC 900
 DB 1073 CGCAGCGCTCGCTCAGGAGACAGCGGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCC 1132
 QY 901 TGCCCGCGGTCGCGGCTTTCGCGCGCTCTTCGCGCGCGCGCGCGCTTCGAGGCGCTGCTG 960
 DB 1133 TGCCCGCGGTCGCGGCTTTCGCGCGCTCTTCGCGCGCGCGCGCGCTTCGAGGCGCTGCTG 1192
 QY 961 CCGCTCTCGCGGTACGCGGCGGCGGCGCGCGCGGCTTTCGCGCGCGCGCGCGCGGAGTG 1020
 DB 1193 CCGCTCTCGCGGTACGCGGCGGCGGCGCGCGGCTTTCGCGCGCGCGCGCGGAGTG 1252
 QY 1021 CCACCGACCGCGCGCGCTTTCGCTTTCGACCTTTCGCGCGCGCGCGCGCGCGCGCGCA 1080
 DB 1253 CCACCGACCGCGCGCGCTTTCGCTTTCGACCTTTCGCGCGCGCGCGCGCGCGCGCGCA 1312
 QY 1081 CTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTTTCGCTTTCGCGCGCGCGCGCGCGCG 1140
 DB 1313 CTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTTTCGCTTTCGCGCGCGCGCGCGCGCG 1372
 QY 1141 CTGAGCGCGCGCTTTCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
 DB 1373 CTGAGCGCGCGCTTTCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
 QY 1201 CTAGCTTGA 1209
 DB 1433 CTAGCTTGA 1441
 RESULT 2
 AL499606 Human DNA sequence from clone RP11-13J16 on chromosome 6, complete
 LOCUS sequence.
 DEFINITION AL499606
 ACCESSION AL499606
 VERSION AL499606.18 GI:17973944
 KEYWORDS HTG
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 62761)
 AUTHORS Corby, N.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 COMMENT On Dec 20, 2001 this sequence version replaced gi:17902504.

QY 1140 CCTGCGAGCGGCT 1153
 Db 2879 CCTGCGAGCGGCT 2892

RESULT 5
 AC067929/c
 LOCUS AC067929.2
 DEFINITION Homo sapiens chromosome 6 clone RP11-403120 map 6, WORKING DRAFT
 SEQUENCE, 18 unordered pieces.
 AC067929
 AC067929.2 GI:8247824
 VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, B., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Malmir, J., Menes, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 182152)

REFERENCE
 AUTHORS Birren, B., Linton, J., Nussbaum, C., Lander, B., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Malmir, J., Menes, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT
 On Jun 4, 2000 this sequence version replaced gi:7655991.
 All repeats were identified using RepeatMasker:

FEATURES
 source
 1. 182152
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6"
 /clone="RP11-403120"

Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L10103
 Center clone name: 403_I_20
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 173395 bases at least Q40
 Consensus quality: 177417 bases at least Q30
 Consensus quality: 179105 bases at least Q20
 Insert size: 179000; agarose-fp
 Insert size: 180452; sum-of-contigs
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 2222: contig of 2222 bp in length
 * 2223: gap of 100 bp
 * 2322: contig of 1463 bp in length
 * 2323: gap of 100 bp
 * 3785: gap of 100 bp
 * 3786: gap of 100 bp
 * 3885: gap of 100 bp
 * 6753: contig of 2868 bp in length
 * 6754: gap of 100 bp
 * 6754: gap of 100 bp
 * 6854: gap of 100 bp
 * 10037: contig of 3183 bp in length
 * 10137: gap of 100 bp
 * 10137: gap of 100 bp
 * 14537: contig of 4401 bp in length
 * 14538: gap of 100 bp
 * 14638: gap of 100 bp
 * 21177: contig of 6540 bp in length
 * 21178: gap of 100 bp
 * 21278: gap of 100 bp
 * 30240: contig of 8963 bp in length
 * 30241: gap of 100 bp
 * 30341: gap of 100 bp
 * 40123: contig of 9783 bp in length
 * 40223: gap of 100 bp
 * 40224: gap of 100 bp
 * 47999: contig of 7776 bp in length
 * 48000: gap of 100 bp
 * 48100: contig of 10136 bp in length
 * 58235: gap of 100 bp
 * 58236: gap of 100 bp
 * 58336: gap of 100 bp
 * 67042: gap of 100 bp
 * 67141: gap of 100 bp
 * 75921: contig of 8780 bp in length
 * 75922: gap of 100 bp
 * 76022: contig of 10506 bp in length
 * 86527: gap of 100 bp
 * 86528: gap of 100 bp
 * 86627: gap of 100 bp
 * 101334: contig of 14707 bp in length
 * 101335: gap of 100 bp
 * 101435: contig of 16196 bp in length
 * 117330: gap of 100 bp
 * 117331: contig of 20563 bp in length
 * 117731: gap of 100 bp
 * 138293: contig of 19448 bp in length
 * 138393: gap of 100 bp
 * 138394: contig of 19448 bp in length
 * 157842: gap of 100 bp
 * 157941: gap of 100 bp
 * 157942: contig of 24211 bp in length.

Db	9676	TTTCTCAGCTCTTTCGCAATCGACAGCATCTCTGGCAAGCCCTTTCGAGCCGCGCTC	9611
QY	856	AGGGACACGGCCCCCGGGACGACGCTTCAGTGGGGCGCGCCGCTGCGCGCGCTGCC	915
Db	9616	AGGGACACGGCCCCCGGGACGACGCTTCAGTGGGGCGCGCCGCTGCGCGCGCTGCC	9557
QY	916	GGGTTCCTCCGCGCTCTCTCCCGCGCGCGCTCTGAGGGCCCTGCTGCGCTCTGCGGGTAC	975
Db	9556	GGGTTCCTCCGCGCTCTCTCCCGCGCGCGCTCTGAGGGCCCTGCTGCGCTCTGCGGGTAC	9497
QY	976	GGCGGGGGGAGCGCGCGCGCTGGCGCGCGCGAGGCGAGTGTCACCGACCGCGCGG	1035
Db	9496	GGCGGGGGGAGCGCGCGCGCTGGCGCGCGCGAGGCGAGTGTCACCGACCGCGCGG	9437
QY	1036	CGCTCTCTCTGTCACCTCTCCGGGGGGCGCGCGCGCAAGCCACTCCGAGGCCCGGG	1095
Db	9436	CGCTCTCTCTGTCACCTCTCCGGGGGGCGCGCGCGCAAGCCACTCCGAGGCCCGGG	9377
QY	1096	GGCGGGGGGCGGACCTGTACTGCCCTCTGCGCTGCCCGCGAGCCCTGCGAGGCGCT	1153
Db	9376	GGCGGGGGGCGGACCTGTACTGCCCTCTGCGCTGCCCGCGAGCCCTGCGAGGCGCT	9319
RESULT 6			
LOCUS	AC012350/c	178168 bp	DNA linear HTG 01-APR-2000
DEFINITION	Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered pieces.		
ACCESSION	AC012350		
VERSION	AC012350.3 GI:7381803		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,B.		
TITLE	Homo sapiens, clone RP11-16N9		
REFERENCE	2 (bases 1 to 178168)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,B., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B., Brown,A., Casle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeAerflinger,K., Dewar,K., Domino,M., Donegan,D., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Apr 1, 2000 this sequence version replaced gi:6479001. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RW/RepeatMasker.html		
	----- Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIBR		
	Web site: http://www-seq.wi.mit.edu		
	Contact: sequence_submissions@genome.wi.mit.edu		
	----- Project Information		
	Center project name: L3593		
	Center clone name: 16 N 9		
	----- Summary Statistics		
	Sequencing vector: M13; W77815; 100% of reads		
	Chemistry: Dye-terminator Big Dye; 100% of reads		
	Assembly program: Phrap; version 0.960731		

Consensus quality: 172070 bases at least Q40
Consensus quality: 174479 bases at least Q30
Consensus quality: 175597 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 176668; sum-of-contigs
Quality coverage: 6.4 in Q20 bases; agarose-fp
Quality coverage: 6.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1694: contig of 1694 bp in length
* 1695
* 1794: gap of 100 bp
* 4437: contig of 2643 bp in length
* 4438
* 4537: gap of 100 bp
* 4538
* 6644: contig of 2107 bp in length
* 6645
* 6646: gap of 100 bp
* 6745
* 10092: contig of 3348 bp in length
* 10093
* 10192: gap of 100 bp
* 10193
* 11801: contig of 1609 bp in length
* 11802
* 11901: gap of 100 bp
* 11902
* 15584: contig of 3683 bp in length
* 15585
* 15685: gap of 100 bp
* 15686
* 19675: contig of 3991 bp in length
* 19676
* 19775: gap of 100 bp
* 19776
* 27913: contig of 8138 bp in length
* 27914
* 28013: gap of 100 bp
* 28014
* 34156: contig of 6143 bp in length
* 34157
* 34256: gap of 100 bp
* 34257
* 43325: contig of 9069 bp in length
* 43326
* 43425: gap of 100 bp
* 43426
* 55265: contig of 11840 bp in length
* 55266
* 55365: gap of 100 bp
* 55366
* 68308: contig of 12941 bp in length
* 68309
* 68408: gap of 100 bp
* 68409
* 82910: contig of 14504 bp in length
* 82911
* 83010: gap of 100 bp
* 83011
* 102181: contig of 19171 bp in length
* 102182
* 102282: gap of 100 bp
* 102283
* 128058: contig of 25777 bp in length
* 128059
* 128158: gap of 100 bp
* 128159
* 178168: contig of 50010 bp in length.

FEATURES

source
1. 178168
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-16N9"
/clone_lib="RP11-16N9 Human Male BAC"
1. 1694
/note="assembly_fragment"
1795. 4437
/note="assembly_fragment"
4538. 6644
/note="assembly_fragment"
6745. 10092
/note="assembly_fragment"
10193. 11801
/note="assembly_fragment"
clone_end:SP6
vector_side:right
11902. 15584
/note="assembly_fragment"
15685. 19675
/note="assembly_fragment"
19776. 27913
/note="assembly_fragment"
28014. 34156

misc_feature /note="assembly_fragment"
34257. 43325
/note="assembly_fragment"
43426. 55265
/note="assembly_fragment"
55366. 68306
/note="assembly_fragment"
68407. 82910
/note="assembly_fragment"
83011. 102181
/note="assembly_fragment"
clone_end:T7
vector_side:right
102282. 128058
/note="assembly_fragment"
128159. 178168
/note="assembly_fragment"

ORIGIN
Query Match 45.8%; Score 554; DB 2; Length 178168;
Best Local Similarity 99.8%; Pred. No. 1.5e-245; Indels 0; Gaps 0;
Matches 604; Conservative 0; Mismatches 1;

QY 549 CTCGCGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTT 608
DB 15584 CTCGCGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGAGTACACCTT 15525
QY 609 GCGGACGGGTCTTCGCGCCCGCCGAGCGCTCAGCCACCGCGCGCGCGCGCGCGCG 668
DB 15524 GCGGACGGGTCTTCGCGCCCGCCGAGCGCTCAGCCACCGCGCGCGCGCGCGCGCG 15465
QY 669 GCGGCGGTGGGGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
DB 15464 GCGGCGGTGGGGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15405
QY 729 GCGCGCGCGCTTCG 788
DB 15404 GCGCGCGCGCTTCG 15345
QY 789 GGGCAAGTTCTCCAGCTCTTCGCGCATCGACAGCATCTCGCGCAAGCGCTTCGCGAGCG 848
DB 15344 GGGCAAGTTCTCCAGCTCTTCGCGCATCGACAGCATCTCGCGCAAGCGCTTCGCGAGCG 15285
QY 849 TGGCTCTAGGACACCG 908
DB 15284 CCGCTCAGGACACCG 15225
QY 909 GCTGCGCGCGTTCCCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 968
DB 15224 GCTGCGCGCGTTCCCGCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 15165
QY 969 CCGGTACCG 1028
DB 15164 CCGGTACCG 15105
QY 1029 CCG 1088
DB 15104 CCG 15045
QY 1089 CCG 1148
DB 15044 CCG 14985
QY 1149 GGCCT 1153
DB 14984 GGCCT 14980

RESULT 7
AC016269/c AC016269 74969 bp DNA linear HTG 13-JUL-2000
LOCUS
DEFINITION Homo sapiens clone RP11-16N9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016269

[illegible]


```
/db_xref="MGI:1298228"
481..1683
/codon_start=1
/product="forkhead box Q1"
/protein_id="AAH47155.1"
/db_xref="GI:28502982"
/db_xref="LocusID:15220"
/translations="MKLEVFVPRAAHGDKMGSDLEGAGSSDVPSPLSAGDDSLGSDG
DCAANSFPAAGSGADLEGGERNSGGPSAQDQPEATDSDRTQASAGPCAGGVGGG
EGARSKYTRRPKPPYSYIALIAMAIRDSAGGLTLAEINELMKGPFPRGSGYTGWR
NSVRHNLSDNCFVKVLRDPSRPMGKNDYMLNPSEYTFADGVFRERKRLSHRTTV
SASGLRPEAPGPAGTPOAPARSSPIARSPARQERSSPASKFSSFAIDSLISK
PFRSRDGSALGVQLPWGAAPCPPLRAYPALLPAAPGALLPLCAYGASEPTLLASR
GTEVQPAAPLLAPLSTAPAKFPRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS
YPVETLLA"
misc_feature
817..1089
/notes="FH; Region: FORKHEAD"
/db_xref="CDD:smart00339"
ORIGIN
Query Match 6.4%; Score 77; DB 10; Length 2046;
Best Local Similarity 99.2%; Pred. No. 1.8e-24;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 526 TTGCTCAAGTGCTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACAATCTGGATGCTC 585
DB 997 TTGCTCAAGTGCTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACAATCTGGATGCTC 1056
QY 586 AACCCCAACAGCAGTACACCTTCGCGCGCGGGGTTCGCGCGCGCGCGCGCGCTC 645
DB 1057 AACCCCAACAGCAGTACACCTTCGCGCGCGGGGTTCGCGCGCGCGCGCGCGCTC 1116
QY 646 AGCCACCG 653
DB 1117 AGCCACCG 1124
RESULT 11
AF010405 4763 bp DNA linear ROD 20-JAN-2000
LOCUS Mus musculus fork head transcription factor (Hfh-1L) gene, complete
DEFINITION
ACCESSION AF010405
VERSION AF010405.2 GI:6716869
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 4763)
AUTHORS Frank, S. and Zoll, B.
TITLE Mouse HNF-3/fork head homolog-1-like gene: structure, chromosomal
location, and expression in adult and embryonic kidney
JOURNAL DNA Cell Biol. 17 (8), 679-688 (1998)
MEDLINE 98392851
PUBMED 9726250
REFERENCE
2 (bases 1 to 4763)
AUTHORS Frank, S. and Zoll, B.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1997) University of Goettingen, Institute of
Human Genetics, Gosslerstr 12d, Goettingen 37073, Germany
REFERENCE
3 (bases 1 to 4763)
AUTHORS Pasche, B., Bieller, A. and Zoll, B.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) University of Goettingen, Institute of
Human Genetics, Heinrich-Dueker-Weg 12, Goettingen 37073, Germany
REMARK Sequence update by submitter
COMMENT On Jan 20, 2000 this sequence version replaced gi:2997586.
FEATURES
source
1..4763
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129"
/db_xref="taxon:10090"
/chromosome="13"
/tissue_type="kidney"
<2467..>3669
/gene="Hfh-1L"
/note="HNF-3/fork head homologue-1 like"
<2467..>3669
/gene="Hfh-1L"
/product="fork head transcription factor"
2467..3669
/note="Hfh-1L"
/feature="member of HNF-3/fork head transcription factor
family"
/codon_start=1
/product="fork head transcription factor"
/protein_id="AAC12973.2"
/db_xref="GI:6716870"
/translations="MKLEVFVPRAAHGDKMGSDLEGAGSSDVPSPLSAGDDSLGSDG
DCAANSFPAAGSGADLEGGERNSGGPSAQDQPEATDSDRTQASAGPCAGGVGGG
EGARSKYTRRPKPPYSYIALIAMAIRDSAGGLTLAEINELMKGPFPRGSGYTGWR
NSVRHNLSDNCFVKVLRDPSRPMGKNDYMLNPSEYTFADGVFRERKRLSHRTTV
SASGLRPEAPGPAGTPOAPARSSPIARSPARQERSSPASKFSSFAIDSLISK
PFRSRDGSALGVQLPWGAAPCPPLRAYPALLPAAPGALLPLCAYGASEPTLLASR
GTEVQPAAPLLAPLSTAPAKFPRGPETAGAAHLYCPLRLPTALQAAAACGPGPHLS
YPVETLLA"
misc_feature
2782..3123
/gene="Hfh-1L"
/note="encodes DNA binding domain"
ORIGIN
Query Match 6.4%; Score 77; DB 10; Length 4763;
Best Local Similarity 99.2%; Pred. No. 1.6e-24;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 526 TTGCTCAAGTGCTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACAATCTGGATGCTC 585
DB 2983 TTGCTCAAGTGCTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACAATCTGGATGCTC 3042
QY 586 AACCCCAACAGCAGTACACCTTCGCGCGCGGGGTTCGCGCGCGCGCGCGCGCTC 645
DB 3043 AACCCCAACAGCAGTACACCTTCGCGCGCGGGGTTCGCGCGCGCGCGCGCGCTC 3102
QY 646 AGCCACCG 653
DB 3103 AGCCACCG 3110
RESULT 12
AL589738 187709 bp DNA linear ROD 29-JUN-2002
LOCUS Mouse DNA sequence from clone RP23-322U11 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL589738
VERSION AL589738.6 GI:15422196
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 187709)
AUTHORS Phillimore, B.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Sep 3, 2001 this sequence version replaced gi:13445573.
COMMENT During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-322J11 is from the RP23-322 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBacE3.6
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: <http://mrcseq.har.mrc.ac.uk>
Contact: moused@har.mrc.ac.uk

FEATURES
source
1. 187709 Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="13"
/clone="RP23-322J11"
/clone_lib="RP23-322J11"

ORIGIN
Query Match 6.4%; Score 77; DB 10; Length 187709;
Best Local Similarity 99.2%; Pred. No. 1.1e-24;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTGCTCAAGTGTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACTACTGGATGCTC 585
Db 136222 TTGCTCAAGTGTGCGGACCCCTCGCGCCCTCGGCGCAAGCAACTACTGGATGCTC 136281

QY 586 AACCCCAACAGCAGTACACTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCCTC 645
Db 136282 AACCCCAACAGCAGTACACTTCGCGCAGCGGGTCTTCGCGCGCGCGCAAGCGCCTC 136341

QY 646 AGCCACCG 653
Db 136342 AGCCACCG 136349

RESULT 13
AF153193
LOCUS AF153193 1576 bp DNA linear ROD 24-AUG-2001
DEFINITION Rattus rattus winged helix/forkhead transcription factor HFH1 (Hfh1) gene, complete cds.
ACCESSION AF153193
VERSION AF153193.1 GI:8132282
KEYWORDS Rattus rattus (black rat)
SOURCE Rattus rattus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1576)
AUTHORS Hong H.-K., Noveroske J.K., Headon D.J., Liu T., Sy M.S., Justice M.J. and Chakravarti A.
TITLE The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice
JOURNAL Genesis 29 (4), 163-171 (2001)
MEDLINE 21207067
PUBMED 11309849
REFERENCE 2 (bases 1 to 1576)
AUTHORS Hong H.-K., Noveroske J.K., Justice M.J. and Chakravarti A.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Genetics, Case Western Reserve University,

10900 Euclid Ave., Cleveland, OH 44106-4955
Location/Qualifiers
1. 1576
/organism="Rattus rattus"
/mol_type="genomic DNA"
/db_xref="taxon:10117"
/genes="Hfh1"
/genes="Hfh1"
/genes="Hfh1"
/product="winged helix/forkhead transcription factor HFH1"
/product="winged helix/forkhead transcription factor HFH1"
/product="winged helix/forkhead transcription factor HFH1"
/protein_id="AA573234.1"
/db_xref="GI:8132283"
/translation="MKLEVPAPRAAHGDMGSLDLEGAGSDVPSLSAAGDDSLGSDG
DGAANGPAAGRGAVDLEGGGERSGGASTQDDPEVTGSRQTQAGVFGPCAGSVGG
EGARSKPYTRRPKPYSYIALIAIRDSAGRLTALINEYLMGKFPFRRGYSYTV
NSVRHLSLNDKCFVKVLRDPSRPMGKDNMNLNPNSEYTPADGVFRRRLSHRTTV
SASGLRPEEAPPGPAGTPAPTAGSSPIARSPAROBEGSSPASKTSSSPFADISLSK
PFRSRDGDPAQVQLPWSAACPPIRAYPALLPSSGGALLPLCAYGAGEPTLLASR
GREVPAPLPLAPLSTAAAPKPFRTAGAAHLICPLRLPTALQAAACGPGPHLS
YRVETLLA"

ORIGIN
Query Match 4.4%; Score 53; DB 10; Length 1576;
Best Local Similarity 100.0%; Pred. No. 2.4e-13;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TACACCTTCGCGACGCGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
Db 686 TACACCTTCGCGACGCGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 738

RESULT 14
RATHFH1
LOCUS RATHFH1 1760 bp mRNA linear ROD 22-AUG-1995
DEFINITION Rattus norvegicus HNF-3/forkhead homolog-1 (HNF-1) mRNA, complete cds.
ACCESSION L13201
VERSION L13201.1 GI:951067
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 1760)
AUTHORS Clevidence D.E., Overdier D.G., Tao W., Qian X., Pani L., Lai E. and Costa R.H.
TITLE Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (9), 3948-3952 (1993)
MEDLINE 93248207
PUBMED 7683413
COMMENT On Aug 22, 1995 this sequence version replaced gi:550512.
Original source text: rat.
Location/Qualifiers
1. 1760
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="lung"
/dev_stage="9 week old"
/tissue_lib="Clontech"
1. 1760
/gene="HNF-1"
250. 1569
/gene="HNF-1"

FEATURES
source
1. 1760
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="lung"
/dev_stage="9 week old"
/tissue_lib="Clontech"
1. 1760
/gene="HNF-1"
250. 1569
/gene="HNF-1"

/note="bp 84. .889 winged-helix DNA binding domain"
/codon_start=1
/product="HNF-3/forhead homolog-1"
/protein_id="AAA74561.1"
/db_xref="GI:550513"
/translation="MGSLEGAGSSDVPSPLSAGDDSLGSDGCAANSPPAAGSGAD
LEGSGENSSGASTQDDPVTGSRTOAGVPQPCAGSVGGEGARSKPVTTRPKPP
YSYIALIMARSDAGRLTAEINELVLMGEPFPGSYTGWRNSVRHNLNDCFVK
VLPDPSRWGNDYMLNPNSEYTPADGVFRRKRRLSHRTVSASVGGSGPPGAG
TPQAPTAGSPAPSPARQEGSPSPKSFADISILKPSRRGRRDGTGLRGCSY
FGALPAPRCAPIPFRFPAPVPCRSVLTRAAHAGVARGAARGAPVARGAPLH
RGPSQAIRSGDRRGRLVPTTAHAGGRLGSGAVPLPGDAASTLGAWSSAG
GSRVELGAPFVMEGTFCGHWAGSRAVHPPSPRNRDFFSTLHQM"

ORIGIN

Query Match 4.4%; Score 53; DB 10; Length 1760;
Best Local Similarity 100.0%; Pred. No. 2.4e-13;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 601 TACACTTCGCGAGCGGGTCTTCGCGCGCCGCGAGCGCTCAGCACCG 653
Db 796 TACACTTCGCGAGCGGGTCTTCGCGCGCCGCGAGCGCTCAGCACCG 848

RESULT 15

AC119497 195184 bp DNA linear HTG 20-NOV-2002
Rattus norvegicus clone CH230-521f8, *** SEQUENCING IN PROGRESS
*** 4 unordered pieces.
AC119497
AC119497.5 GI:25137938
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 195184)
Muzny, D.Marie, Metzker, M.Jee., Abruzzo, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chu, J.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, V., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, H., Loulesed, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, B.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokeleneh, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Poppert, F., Poudexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Taylor, C.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmari, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 195184)
Worley, K.C.
Direct Submission
Submitted (27-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 195184)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23195197.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWQI
Center clone name: CH230-521f8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 158913 bases at least Q40
Consensus quality: 162109 bases at least Q30
Consensus quality: 164133 bases at least Q20
Estimated insert size: 161675; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 16506: contig of 16506 bp in length
* 16507 180638: gap of unknown length
* 16607 180639: contig of 164032 bp in length
* 180639 180738: gap of unknown length
* 180739 182447: contig of 1709 bp in length
* 182448 182547: gap of unknown length

ORIGIN

Search completed: March 31, 2004, 03:52:56
Job time : 4683 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 01:11:33 ; Search time 504 Seconds
(without alignments)
10190.618 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggagggtgttgcgt.....tggagacgctcctagcttga 1209

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002s:*
- 7: Geneseq2003s:*
- 8: Geneseq2003bs:*
- 9: Geneseq2003cs:*
- 10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1209	100.0	1209	6 ABQ81557	Abq81557 Gene up-r
2	53	4.4	1760	7 ABT42112	Abt42112 Toxicity
3	37	3.1	658	6 ABQ39023	Abq39023 Oligonuc1
4	37	3.1	658	6 ABQ39022	Abq39022 Oligonuc1
5	30	2.5	1387	5 AAS72684	Aas72684 DNA encod
6	30	2.5	3786	5 AAS72683	Aas72683 DNA encod
7	27	2.2	514	6 ABQ34097	Abq34097 Oligonuc1
8	27	2.2	514	6 ABQ34096	Abq34096 Oligonuc1
9	27	2.2	741	6 ABQ34746	Abq34746 Oligonuc1
10	27	2.2	741	6 ABQ34747	Abq34747 Oligonuc1
11	27	2.2	3482	7 ABX76288	Abx76288 Lung canc
12	26	2.2	738	6 ABQ25892	Abq25892 Oligonuc1
13	26	2.2	738	6 ABQ25893	Abq25893 Oligonuc1
14	25	2.1	441	6 ABT80120	Abt80120 Human ova
15	25	2.1	890	6 ABQ34485	Abq34485 Oligonuc1
16	25	2.1	890	6 ABQ34484	Abq34484 Oligonuc1
17	25	2.1	1965	9 ABQ31305	Abq31305 Testoster
18	25	2.1	2187	7 ABQ34783	Abq34783 Coding se
19	25	2.1	4895	2 AAQ79378	Aaq79378 Human N-m
20	25	2.1	4895	2 AAQ72917	Aaq72917 Human N-m
21	25	2.1	4895	3 AAQ38731	Aaq38731 Human NMD
22	25	2.1	4895	3 AAA95040	Aaa95040 Human N-m
23	25	2.1	4895	6 ABI99196	Abi99196 Human NMD

24	25	2.1	4695	6 AAL47384	Aal47384 Human NMD
25	25	2.1	4695	7 ABX98543	Abx98543 Human N-m
26	25	2.1	4695	7 ABX92892	Abx92892 Human N-m
27	25	2.1	4695	7 ABX77688	Abx77688 DNA encod
28	25	2.1	4695	8 ACD98353	AcD98353 Human NMD
29	25	2.1	4695	9 AAD59685	Aad59685 Human NMD
30	24	2.0	51	4 AAL29440	Aal29440 Human SNP
31	24	2.0	146	3 AAC09019	Aac09019 Human sec
32	24	2.0	159	4 AAH27142	Aah27142 Human cyc
33	24	2.0	310	6 AAD37498	Aad37498 COX2 prom
34	24	2.0	429	3 AAC57690	Aac57690 Arachidon
35	24	2.0	1056	6 ABZ35308	Abz35308 Human gen
36	24	2.0	1782	8 ACC85255	Acc85255 Craniofac
37	24	2.0	1860	2 AAQ50632	Aaq50632 Brain fac
38	24	2.0	1979	4 ABL58042	AbL58042 Human cyc
39	24	2.0	2271	6 ABZ35021	Abz35021 Human gen
40	24	2.0	2487	9 ADB75311	Adb75311 Prostate
41	24	2.0	2563	7 ABX63477	Abx63477 Human CDN
42	24	2.0	3387	2 AAQ71002	Aaq71002 Cyclooxyg
43	24	2.0	3387	2 AAQ89376	Aaq89376 Human cyc
44	24	2.0	3387	3 AAA34993	Aaa34993 Human ade
45	24	2.0	3387	3 AAF21115	Aaf21115 Human low

ALIGNMENTS

RESULT 1

ABQ81557 ID ABQ81557 standard; cDNA; 1209 BP.

XX AC ABQ81557;

DT 30-DEC-2002 (first entry)

DE Gene up-regulated in metastatic colorectal cancer.

XX Colorectal cancer; metastasis; differential expression; cytostatic;
diagnosis; gene therapy; vaccine; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1209

FT /*tag= a

FT /product= "Metastatic colon cancer polypeptide"

XX PN WO200268677-A2.

XX PD 06-SEP-2002.

XX 27-FEB-2002; 2002WO-US006001.

XX 27-FEB-2001; 2001US-0272206P.

PR 02-APR-2001; 2001US-0281149P.

PR 17-APR-2001; 2001US-0284555P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

PA (UYCA-) UNIV CASE WESTERN RESERVE.

XX Mack DH, Markowitz SD;

XX WPI; 2002-698677/75.

DR P-FSDB; ABP54691.

XX New genes that are up- or down-regulated in colorectal cancer, useful for
diagnosing colorectal cancer in a subject, or for identifying modulators
of colorectal cancer-associated proteins and genes for treating
colorectal cancer.

XX Claim 5; Page 252; 260pp; English.

XX The present sequence is the nucleotide sequence of a human gene that

CC exhibits increased expression in metastatic colorectal cancer (MCC)
 CC samples. The gene is up-regulated in colon cancer-derived metastases
 CC compared to normal colon tissue. It is an example of claimed nucleic acid
 CC molecules that are up- or down-regulated in metastatic colorectal cancer
 CC cells. Such MCC-associated nucleic acids are useful in diagnostic and
 CC prognostic applications, in screening applications e.g. biochips, for
 CC identification of variant MCC-associated sequences, in informatics, for
 CC expression of MCC-associated proteins, in drug screening assays for
 CC identification of modulators of MCC-associated proteins, which are useful
 CC for treating cancer, in gene therapy, as DNA vaccines, and as
 CC antisense or ribozyme modulators of MCC

XX Sequence 1209 BP; 150 A; 485 C; 432 G; 141 T; 0 U; 1 Other;

Query Match 100.0%; Score 1209; DB 6; Length 1209;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTTGAGGTGTTCTGCTCCCTCGCGGCGCCACAGGGGACAGAGGCGAGTGACCTG 60
 DB 1 ATGAAGTTGAGGTGTTCTGCTCCCTCGCGGCGCCACAGGGGACAGAGGCGAGTGACCTG 60
 QY 61 GAGGCGCGGCGGCGAGGCGGCGGCTCCCGCTGTCCGCGGCGGAGGAGGAGTCCCTG 120
 DB 61 GAGGCGCGGCGGCGAGGCGGCGGCTCCCGCTGTCCGCGGCGGAGGAGGAGTCCCTG 120
 QY 121 GGCTCAGATGGGACTGCGCGGCGAAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
 DB 121 GGCTCAGATGGGACTGCGCGGCGAAGCGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
 QY 181 GCGGACGGCGAAGCAGAGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240
 DB 181 GCGGACGGCGAAGCAGAGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 240
 QY 241 GCTGCTGACGCGTGTGTCGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
 DB 241 GCTGCTGACGCGTGTGTCGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
 QY 301 GCGCGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360
 DB 301 GCGCGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360
 QY 361 TACTCGTACATCGCGTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420
 DB 361 TACTCGTACATCGCGTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420
 QY 421 CTGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGGCGGCGGCGGCGGCGG 480
 DB 421 CTGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTCTTTTTCGCGGCGGCGGCGGCGGCGG 480
 QY 481 GGCTGCGCAACTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540
 DB 481 GGCTGCGCAACTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540
 QY 541 CGGGAACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 DB 541 CGGGAACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 QY 601 TACACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 DB 601 TACACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 QY 661 GTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 DB 661 GTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 QY 721 GCGCGCG 780
 DB 721 GCGCGCG 780
 QY 781 AGCCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 DB 781 AGCCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840

QY 841 CGCAGCGTGCCTCAGGACACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 DB 841 CGCAGCGTGCCTCAGGACACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 QY 901 TGCCCGCGGCGTGCCTCAGGACACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 DB 901 TGCCCGCGGCGTGCCTCAGGACACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 QY 961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 DB 961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 QY 1021 CCACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 DB 1021 CCACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 QY 1081 CTCGAGCGG 1140
 DB 1081 CTCGAGCGG 1140
 QY 1141 CTGAGGCGG 1200
 DB 1141 CTGAGGCGG 1200
 QY 1201 CTAGCTTGA 1209
 DB 1201 CTAGCTTGA 1209

RESULT 2
 ABT42112
 ID ABT42112 standard; DNA; 1760 BP.
 XX ABT42112;
 DT 26-JUN-2003 (first entry)
 DE Toxicity modelling related rat gene SEQ ID No 1814.
 DE Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 KW database; drug screening; toxicity assay; rat; ds.
 OS Rattus norvegicus.
 PN WO200295000-A2.
 XX 28-NOV-2002.
 PF 22-MAY-2002; 2002WO-US016173.
 XX 22-MAY-2001; 2001US-0292335P.
 PR 13-JUN-2001; 2001US-0297523P.
 PR 19-JUN-2001; 2001US-0298925P.
 PR 10-JUL-2001; 2001US-0303807P.
 PR 10-JUL-2001; 2001US-0303808P.
 PR 28-AUG-2001; 2001US-0315047P.
 PR 27-SEP-2001; 2001US-0324928P.
 PR 22-OCT-2001; 2001US-0330462P.
 PR 01-NOV-2001; 2001US-0330867P.
 PR 21-NOV-2001; 2001US-0331805P.
 PR 06-DEC-2001; 2001US-0336144P.
 PR 19-DEC-2001; 2001US-0340873P.
 PR 21-FEB-2002; 2002US-0357843P.
 PR 21-FEB-2002; 2002US-0357843P.
 PR 15-MAR-2002; 2002US-0364134P.
 PR 08-APR-2002; 2002US-0370144P.
 PR 08-APR-2002; 2002US-0370266P.
 PR 17-APR-2002; 2002US-0370247P.
 PR 21-APR-2002; 2002US-0372794P.
 PR 21-APR-2002; 2002US-0371679P.

```

XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
XX modeling, comprises preparing a gene expression profile of a tissue or
XX cell sample exposed to the compound, and comparing the gene expression
XX profile to a database.
XX
XX Example 4; Page; 446pp; English.
XX
XX The invention relates to a novel method of predicting at least one toxic
XX effect of a compound. The method comprises a gene expression profile of a
XX tissue or cell sample exposed to the compound, and comparing the gene
XX expression profile to a database comprising at least part of the data or
XX information given in the specification. The methods are useful for
XX predicting at least one toxic effect of a compound, predicting the
XX progression of a toxic effect of a compound, predicting the renal
XX toxicity of a compound, or identifying toxicity markers in tissues or
XX cells exposed to known renal toxin. The genes are useful as toxicity
XX markers in drug screening and toxicity assays, in monitoring disease or
XX physiological states, or disease progression. This polynucleotide
XX represents a rat DNA sequence relating to the toxic effect database
XX described in the specification. NOTE: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from the World Intellectual Property
XX Organization
XX
XX Sequence 1760 BP; 301 A; 615 C; 535 G; 309 T; 0 U; 0 Other;
XX
XX Query Match 4.4%; Score 53; DB 7; Length 1760;
XX Best Local Similarity 100.0%; Pred. No. 1e-13;
XX Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 601 TACACCTTCGCGACGGGTCTTCGCGCGCGCGCGCGCGCTCAGCCACCG 653
XX Db 796 TACACCTTCGCGACGGGTCTTCGCGCGCGCGCGCGCGCTCAGCCACCG 848
XX
XX RESULT 3
XX ABQ39023
XX ID ABQ39023 standard; DNA; 658 BP.
XX
XX AC ABQ39023;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25614.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PP 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX

```

```

XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX
XX Sequence 658 BP; 173 A; 310 C; 108 G; 67 T; 0 U; 0 Other;
XX
XX Query Match 3.1%; Score 37; DB 6; Length 658;
XX Best Local Similarity 100.0%; Pred. No. 1.6e-06;
XX Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 699 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
XX Db 102 CTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138
XX
XX RESULT 4
XX ABQ39022/C
XX ID ABQ39022 standard; DNA; 658 BP.
XX
XX AC ABQ39022;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 25613.
XX
XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PP 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX

```


sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 2.5%; Score 30; DB 5; Length 3786;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGGCTCATC 381
DB 322 AAGCCCCCTACTCGTACATCGGCTCATC 351

RESULT 7

ABQ34097
ID ABQ34097 standard; DNA; 514 BP.

AC ABQ34097;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20688.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

PN 07-MAR-2002.

XX 01-SEP-2001; 2001WO-RP010074.

XX 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert

cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention

Query Match 2.2%; Score 27; DB 6; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CCCCCCTACTCGTACATCGGCTCATC 381

DB 138 CCCCCCTACTCGTACATCGGCTCATC 164

RESULT 8

ABQ34096/C

ID ABQ34096 standard; DNA; 514 BP.

XX ABQ34096;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 20687.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-RP010074.

XX 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member,

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX Sequence 514 BP; 78 A; 42 C; 205 G; 189 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 27; DB 6; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 355 CCCCCCTACTCGTACATCGCGCTCATC 381
DB 377 CCCCCCTACTCGTACATCGCGCTCATC 351
RESULT 9
ID ABQ34746 standard; DNA; 741 BP.
XX
AC ABQ34746;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21337.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC the amplicon.

CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX Sequence 741 BP; 112 A; 62 C; 259 G; 308 T; 0 U; 0 Other;
SQ
Query Match 2.2%; Score 27; DB 6; Length 741;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 355 CCCCCCTACTCGTACATCGCGCTCATC 381
DB 285 CCCCCCTACTCGTACATCGCGCTCATC 259
RESULT 10
ID ABQ34747 standard; DNA; 741 BP.
XX
AC ABQ34747;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21338.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC the amplicon, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

XX SQ Sequence 741 BP; 308 A; 259 C; 62 G; 112 T; 0 U; 0 Other;
 XX
 XX Query Match 2.2%; Score 27; DB 6; Length 741;
 XX Best Local Similarity 100.0%; Pred. No. 0.048; 0; Indels 0; Gaps 0;
 XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 CCCCCCTACTCGTACATCGCGTCTATC 381
 DB 457 CCCCCCTACTCGTACATCGCGTCTATC 483

RESULT 11
 ABX76288
 ID ABX76288 standard; DNA; 3482 BP.
 XX AC ABX76288;
 XX
 XX 02-APR-2003 (first entry)
 XX
 XX Lung cancer-associated polynucleotide #152.
 XX
 XX Lung cancer-associated polynucleotide; gene; ds, cytotatic; emphysema;
 XX antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 XX Unidentified.
 XX
 XX WO200286443-A2.
 XX
 XX 31-OCT-2002.
 XX
 XX 18-APR-2002; 2002WO-US012476.
 XX
 XX 18-APR-2001; 2001US-0284770P.
 XX 10-MAY-2001; 2001US-0290492P.
 XX 09-NOV-2001; 2001US-0339245P.
 XX 13-NOV-2001; 2001US-0350666P.
 XX 29-NOV-2001; 2001US-0334370P.
 XX 12-APR-2002; 2002US-037246P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 XX Aziz N, Murray R;
 XX
 XX WPI; 2003-093161/08.
 XX P-PSDB; ABUS6559.
 XX
 XX Detecting a lung cancer-associated transcript in a cell from a patient
 XX for treating lung cancer, by contacting a biological sample from the
 XX patient with a polynucleotide that exhibits increased or decreased
 XX expression in lung cancer.
 XX
 XX Claim 22; Page 303-304; 453pp; English.
 XX
 XX The invention relates to a method for detecting a lung cancer-associated
 XX transcript in a cell from a patient, comprising contacting a biological
 XX sample from the patient with a polynucleotide that selectively hybridises
 XX to a sequence that is at least 80 % identical to a gene that exhibits
 XX increased or decreased expression in lung cancer samples. Lung cancer-
 XX associated polynucleotides and polypeptides are used for identifying a
 XX compound that modulates a lung cancer-associated polypeptide, for
 XX inhibiting proliferation of a lung cancer-associated cell to treat lung

CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABX76124-ABX76474 represent lung cancer-associated polynucleotides of the
 CC invention

XX SQ Sequence 3482 BP; 716 A; 1137 C; 981 G; 647 T; 0 U; 1 Other;
 XX
 XX Query Match 2.2%; Score 27; DB 7; Length 3482;
 XX Best Local Similarity 100.0%; Pred. No. 0.042; 0; Indels 0; Gaps 0;
 XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 TACATCGGCTCATCGCCATGGCCATC 393
 DB 832 TACATCGGCTCATCGCCATGGCCATC 858

RESULT 12
 ABQ25892/c
 ID ABQ25892 standard; DNA; 738 BP.
 XX AC ABQ25892;
 XX
 XX 12-JUL-2002 (first entry)
 XX
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 12483.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX
 XX Homo sapiens.
 XX
 XX WO200218632-A2.
 XX
 XX 07-MAR-2002.
 XX
 XX 01-SEP-2001; 2001WO-EP010074.
 XX
 XX 01-SEP-2000; 2000DE-01043826.
 XX 05-SEP-2000; 2000DE-01044543.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 XX WPI; 2002-371829/40.
 XX
 XX Determining the degree of cytosine methylation in genomic DNA, useful for
 XX diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 XX genomic sample of DNA. The sample is treated chemically to convert
 XX cytosine (C) but not methylated C, to uracil, then part of the genomic
 XX DNA that contains the target C is amplified to form a labeled amplicon.
 XX The amplicon is hybridised to two classes, each with at least one member,
 XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 XX degree of hybridisation to both classes is determined from the label on
 XX the amplicon. From the ratio of labels hybridised to the two classes of
 XX oligomers, the degree of methylation is calculated. The method is used:
 XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 XX and of a wide range of diseases, e.g. cancer, disorders of the central

CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ3410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

CC Sequence 738 BP; 106 A; 63 C; 274 G; 295 T; 0 U; 0 Other;
 SQ

Query Match 2.2%; Score 26; DB 6; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 CCCCTACTCGTACATCGGCTCATC 381
 Db 373 CCCCTACTCGTACATCGGCTCATC 348

RESULT 13
 ABQ25893
 ID ABQ25893 standard; DNA; 738 BP.
 XX
 AC ABQ25893;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 12484.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 FN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX

This invention describes a novel method for determining the degree of
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 genomic sample of DNA. The sample is treated chemically to convert
 cytosine (C) but not methylated C, to uracil, then part of the genomic
 DNA that contains the target C is amplified to form a labeled amplicon.
 The amplicon is hybridised to two classes, each with at least one member,
 of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 degree of hybridisation to both classes is determined from the label on
 the amplicon. From the ratio of labels hybridised to the two classes of
 oligomers, the degree of methylation is calculated. The method is used:
 (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 and of a wide range of diseases, e.g. cancer, disorders of the central
 nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for

CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ3410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention

CC Sequence 738 BP; 295 A; 274 C; 63 G; 106 T; 0 U; 0 Other;
 SQ

Query Match 2.2%; Score 26; DB 6; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 CCCCTACTCGTACATCGGCTCATC 381
 Db 366 CCCCTACTCGTACATCGGCTCATC 391

RESULT 14
 ABL80120
 ID ABL80120 standard; cDNA; 441 BP.
 XX
 AC ABL80120;
 XX
 DT 17-MAY-2002 (first entry)
 XX
 DE Human ovarian cancer related cDNA clone SEQ ID NO:3098.
 XX
 KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200192581-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US017756.
 XX
 PR 26-MAY-2000; 2000US-0207484P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Algate PA, Harlocker SL, Jones R;
 XX
 DR WPI; 2002-122075/16.
 XX
 PT Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide.
 XX
 PS Claim 1; SEQ ID NO 3098; 489pp; English.
 XX

The present invention describes a composition (i) comprising: carriers
 and immunostimulants; and a polypeptide (ii) of a ovarian tumour
 polypeptide encoded by a polynucleotide (iii) having a cDNA sequence (S1)
 from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 (iii) encoding (ii) having a sequence (S2), a T cell population of (ii),
 or antigen presenting cells that express (ii). (i) has cytostatic
 activity. An oligonucleotide (iv) that hybridises to (S1) can be used for
 detecting ovarian cancer in a patient's biological sample preferably
 serum or ovarian tissue. The method comprises contacting a biological
 sample from a patient with (iv), detecting the amount of polynucleotide
 hybridising to (iv) and comparing the amount to a predetermined cutoff
 value and thereby detecting ovarian cancer in the patient, where the
 amount of polynucleotide hybridising to (iv) is detected preferably by
 polymerase chain reaction (PCR). (i) comprising (iii) and/or (ii) is
 useful for stimulating and/or expanding T cells specific for an ovarian
 tumour protein comprising contacting T cells with (iii) or (ii). (iii) is
 useful in design and preparation of ribozyme molecules for inhibiting
 expression of the tumour polypeptides and proteins in tumour cells; and
 to isolate a full length gene from a suitable library e.g., a tumour cDNA
 library using well known techniques

SQ Sequence 441 BP; 95 A; 116 C; 96 G; 134 T; 0 U; 0 Other;
Query Match 2.1%; Score 25; DB 6; Length 441;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1181 CGTACCGGTGGAGAGCTCTTAGC 1205
Db 13 CGTACCGGTGGAGAGCTCTTAGC 37

Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 358 CCTACTCGTACATCGGCTCATCG 382
Db 59 CCTACTCGTACATCGGCTCATCG 83
Search completed: March 31, 2004, 02:34:43
Job time : 507 secs

RESULT 15
ABQ34485
ID ABQ34485 standard; DNA; 890 BP.
XX AC ABQ34485;
XX AC ABQ34485;
DT 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 21076.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EF010074.
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ3410-
CC ABQ34121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX Sequence 890 BP; 311 A; 354 C; 106 G; 119 T; 0 U; 0 Other;

Query Match 2.1%; Score 25; DB 6; Length 890;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 02:18:33 ; Search time 114 Seconds
(without alignments)
5885.400 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209
Sequence: 1 atgaagtggagggttcgt.....tggagagctcctagcttga 1209

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	6.6	320	4	US-09-833-381-1266
2	25	2.1	1965	4	US-09-220-132-21
3	25	2.1	4695	2	US-08-231-193A-57
4	25	2.1	4695	2	US-08-486-293A-57
5	25	2.1	4695	3	US-08-940-086A-57
6	25	2.1	4695	4	US-08-940-035A-57
7	25	2.1	4695	4	US-08-935-105A-57
8	25	2.1	4695	4	US-09-648-797-57
9	25	2.1	4695	4	US-09-386-123-57
10	24	2.0	159	3	US-09-437-457-11
11	24	2.0	429	4	US-09-641-638-324
12	24	2.0	1860	2	US-08-331-644-3
13	24	2.0	1860	5	PCT-US93-04102-3
14	24	2.0	3387	1	US-08-064-271-11
15	24	2.0	3387	2	US-08-627-254C-29
16	24	2.0	3387	4	US-09-023-655-1050
17	22	1.8	3387	3	US-09-530-589A-19
18	22	1.8	3387	4	US-09-599-781-19
19	20	1.7	480	4	US-09-674-608A-4
20	20	1.7	1454	4	US-09-372-422A-19
21	20	1.7	2561	4	US-09-616-289-48
22	20	1.7	4403765	3	US-09-103-840A-2
23	20	1.7	4411529	3	US-09-103-840A-1
24	19	1.6	576	4	US-09-252-991A-13112
25	19	1.6	576	4	US-09-252-991A-12687
26	19	1.6	720	4	US-09-252-991A-2351
27	19	1.6	720	4	US-09-252-991A-2640

C	28	19	1.6	1101	4	US-09-489-039A-2638	Sequence 2638, Ap
	29	19	1.6	1192	3	US-09-318-448-18	Sequence 18, Appl
C	30	19	1.6	1248	4	US-09-489-039A-6144	Sequence 6144, Ap
	31	19	1.6	1278	4	US-09-252-991A-2483	Sequence 2483, Ap
	32	19	1.6	1563	4	US-09-252-991A-2566	Sequence 2566, Ap
	33	19	1.6	2124	4	US-09-252-991A-12830	Sequence 12830, A
	34	19	1.6	2634	3	US-08-941-936-1	Sequence 1, Appli
C	35	19	1.6	3623	1	US-08-306-691B-35	Sequence 1, Appli
	36	19	1.6	43280	2	US-08-804-227C-1	Sequence 1, Appli
C	37	18	1.5	18	2	US-08-627-254C-12	Sequence 12, Appl
	38	18	1.5	248	4	US-09-397-787-24	Sequence 24, Appl
	39	18	1.5	420	4	US-09-252-991A-7967	Sequence 7967, Ap
	40	18	1.5	504	4	US-09-252-991A-9584	Sequence 9584, Ap
	41	18	1.5	957	4	US-09-252-991A-51	Sequence 51, Appl
	42	18	1.5	1077	4	US-09-522-714-17	Sequence 17, Appl
C	43	18	1.5	1143	4	US-09-252-991A-5748	Sequence 5748, Ap
	44	18	1.5	1188	4	US-09-252-991A-5131	Sequence 5131, Ap
C	45	18	1.5	1239	4	US-09-252-991A-9666	Sequence 9666, Ap

ALIGNMENTS

RESULT 1
US-09-833-381-1266
; Sequence 1266, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1266
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(320)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1266

Query Match 6.6%; Score 80; DB 4; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.4e-24;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 567 GGCACTACTGATGCTCAACCCCAACGCGAGTACACCTTCGCGACGGGTCTTCG 626
Db 1 GGCACTACTGATGCTCAACCCCAACGCGAGTACACCTTCGCGACGGGTCTTCG 60
Qy 627 CCGCGCGCGCAAGCGCTCA 646
Db 61 CCGCGCGCGCAAGCGCTCA 80

RESULT 2
US-09-220-132-21
; Sequence 21, Application US/09220132
; Patent No. 6505607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25

;; PRIOR APPLICATION NUMBER: US 60/068,821
;; PRIOR FILING DATE: 1997-12-24
;; NUMBER OF SEQ ID NOS: 191
;; SOFTWARE: Fast-Seq for Windows Version 4.0
;; SEQ ID NO 21
;; LENGTH: 1965
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-220-132-21

Query Match 2.1%; Score 25; DB 4; Length 1965;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTCGTACATCGCGCTCATCG 382
DB 196 CCCTACTCGTACATCGCGCTCATCG 220

RESULT 3
US-08-231-193A-57
; Sequence 57, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231.193A
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,459
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4595 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
US-08-231-193A-57

Query Match 2.1%; Score 25; DB 2; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-231-193A-57
; Sequence 57, Application US/08231193A
; Patent No. 5849895
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.

QY 285 GGGGCCAGCGCGCGCGCGGGG 309
DB 3700 GGGGCCAGCGCGCGCGGGG 3724

RESULT 4
US-08-486-273A-57
; Sequence 57, Application US/08486273A
; Patent No. 5985586
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,273A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
US-08-486-273A-57

Query Match 2.1%; Score 25; DB 2; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGGG 309
DB 3700 GGGGCCAGCGCGCGGGG 3724

RESULT 5
US-08-940-086A-57
; Sequence 57, Application US/08940086A
; Patent No. 611091
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.

US-08-940-086A-57
; Sequence 57, Application US/08940086A
; Patent No. 611091
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.

```

; APPLICANT: Liaw, Chen W.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,086A
; FILING DATE: 29-SEPT-97
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24735-9383C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 450-8400
; TELEFAX: (619) 450-8499
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
; US-08-940-086A-57

Query Match 2.1%; Score 25; DB 3; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGGCGCGGCGGCGGCGG 309
Db 3700 GGGGCCAGGCGGCGGCGGCGG 3724

RESULT 6
US-08-940-035A-57
; Sequence 57, Application US/08940035A
; Patent No. 6316611
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,105A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:

```

```

; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,035A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9383E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
; US-08-940-035A-57

Query Match 2.1%; Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGGCGGCGGCGGCGGCGG 309
Db 3700 GGGGCCAGGCGGCGGCGGCGGCGG 3724

RESULT 7
US-08-935-105A-57
; Sequence 57, Application US/08935105A
; Patent No. 6376660
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrie P.
; APPLICANT: Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,105A
; FILING DATE: 29-SEPT-97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/231,193
; FILING DATE: 20-APR-1994
; PRIOR APPLICATION DATA:

```



```
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 6362-9383D
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 619-238-0999
/ TELEFAX: 619-238-0062
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: CDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 485..4495
/ US-08-935-105A-57

Query Match      2.1% Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGCGGG 309
Db 3700 GGGGCCAGCGCGCGCGCGGGG 3724

RESULT 8
US-09-648-797-57
/ Sequence 57, Application US/09648797
/ Patent No. 6469142
/ GENERAL INFORMATION:
/ APPLICANT: Daggett, Lorrie P.
/ APPLICANT: Ellis, Steven B.
/ APPLICANT: Liaw, Chen W.
/ APPLICANT: Lu, Chin-Chun
/ TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
/ NUMBER OF SEQUENCES: 63
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Heller Ehrman White & McAuliffe
/ STREET: 4250 Executive Square, 7th Floor
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/648,797
/ FILING DATE: 28-Aug-2000
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/940,086A
/ FILING DATE: 29-SEPT-97
/ APPLICATION NUMBER: US 08/231,193
/ FILING DATE: 20-APR-1994
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 24735-9383C
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 450-8400
/ TELEFAX: (619) 450-8499
/ INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:
LENGTH: 4695 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 485..4495
US-09-648-797-57

Query Match      2.1% Score 25; DB 4; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGGG 309
Db 3700 GGGGCCAGCGCGCGCGGGG 3724

RESULT 9
US-09-386-123-57
/ Sequence 57, Application US/09386123
/ Patent No. 6521413
/ GENERAL INFORMATION:
/ APPLICANT: Daggett, Lorrie P.
/ APPLICANT: Ellis, Steven B.
/ APPLICANT: Liaw, Chen W.
/ APPLICANT: Lu, Chin-Chun
/ TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR
/ NUMBER OF SEQUENCES: 63
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Heller Ehrman White & McAuliffe
/ STREET: 4250 Executive Square, 7th Floor
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: U.S.A.
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/386,123
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/486,273
/ FILING DATE: 06-JUNE-95
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/231,193
/ FILING DATE: 20-APR-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/052,449
/ FILING DATE: 20-APR-1993
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 6362-9383F
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 858-450-8403
/ TELEFAX: 858-587-5360
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4695 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: both
/ TOPOLOGY: both
/ MOLECULE TYPE: CDNA
/ FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 485..4495
US-09-386-123-57

Query Match
Best Local Similarity 100.0%; DB 4; Length 4695;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGGGCGGGCGGGCGGG 309
Db 3700 GGGGCCAGCGCGGGCGGGCGGGCGGG 3724

RESULT 10
US-09-437-457-11
; Sequence 11, Application US/09437457
; Patent No. 6273893
; GENERAL INFORMATION:
; APPLICANT: Giordano, Anthony
; APPLICANT: Xavier, Ashish
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES AND METHODS FOR
; TITLE OF INVENTION: IDENTIFYING COMPOUNDS THAT AFFECT RNA/RNA BINDING PROTEIN
; TITLE OF INVENTION: INTERACTIONS AND MRNA FUNCTIONALITY
; FILE REFERENCE: 50093/014001
; CURRENT APPLICATION NUMBER: US/09/437,457
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-437-457-11

Query Match
Best Local Similarity 100.0%; DB 3; Length 159;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GCGCGCGCGCTCGCGCGCGCTGC 913
Db 73 GCGCGCGCGCTCGCGCGCGCTGC 96

RESULT 11
US-09-641-638-324
; Sequence 324, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 324
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
```

```
; LOCATION: 386
; OTHER INFORMATION: 10-52-386 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 366..385
; OTHER INFORMATION: 10-52-386.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 387..406
; OTHER INFORMATION: 10-52-386.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..18
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 412..429
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 374..398
; OTHER INFORMATION: 10-52-386 potential probe
US-09-641-638-324

Query Match
Best Local Similarity 100.0%; DB 4; Length 429;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GCGCGCGCGCTCGCGCGCGCTGC 913
Db 160 GCGCGCGCGCTCGCGCGCGCTGC 183

RESULT 12
US-08-331-644-3
; Sequence 3, Application US/08331644
; Patent No. 5976872
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,644
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1860 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-331-644-3
```

Query Match 2.0%; Score 24; DB 2; Length 1860;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTCGTACATCGCGCTCATC 381
DB 759 CCCTACTCGTACATCGCGCTCATC 782

RESULT 13

PCT-US93-04102-3
; Sequence 3, Application PC/TUS9304102
; GENERAL INFORMATION:
; APPLICANT: Tao, Wufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04102
; FILING DATE: 19930430
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,292
; FILING DATE: 13-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-977-9550
; TELEFAX: 212-664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1860 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
PCT-US93-04102-3

Query Match 2.0%; Score 24; DB 5; Length 1860;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 CCCTACTCGTACATCGCGCTCATC 381
DB 759 CCCTACTCGTACATCGCGCTCATC 782

RESULT 14

US-08-084-271-11
; Sequence 11, Application US/08064271
; Patent No. 5543297
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Brian P.
; APPLICANT: Cromlish, Wanda A.
; APPLICANT: Mancini, Joseph A.
; APPLICANT: O'Neill, Gary
; APPLICANT: Vickers, Phillip J.

; APPLICANT: Wong, Elizabeth
; TITLE OF INVENTION: HUMAN CYCLOOXYGENASE-2 CDNA AND
; TITLE OF INVENTION: ASSAY FOR EVALUATING CYCLOOXYGENASE ACTIVITY
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Avenue
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.4kb
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7
; SOFTWARE: Microsoft Word 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/064,271
; FILING DATE: 19930506
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Panzer, Curtis C.
; REGISTRATION NUMBER: 33,752
; REFERENCE/DOCKET NUMBER: 189061A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3199
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3387 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-064-271-11

Query Match 2.0%; Score 24; DB 1; Length 3387;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 890 GGGCGCGCCCTGCGCGCGGTGC 913
DB 73 GGGCGCGCCCTGCGCGCGGTGC 96

RESULT 15

US-08-627-254C-29
; Sequence 29, Application US/08627254C
; Patent No. 5859229
; GENERAL INFORMATION:
; APPLICANT: Kniss, Douglas A.
; TITLE OF INVENTION: Eicosanoid Formation
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter & Griwold LLP
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/627,254C
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Goltick, Mary E
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 18525/00107

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3387 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-627-254C-29

Query Match      2.0%; Score 24; DB 2; Length 3387;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      890 GCGCGCGCGCTGCGCGCGCTGC 913
Db      73  GCGCGCGCGCTGCGCGCGCTGC 96

Search completed: March 31, 2004, 04:45:16
JOB time : 132 secs
```



```
181 GGCACGGCGAAACAGAGTGGGAGGCGGGCGGGCGGAGGAGCGGATCCGGGACGA 240
181 GGCACGGCGAAACAGAGTGGGAGGCGGGCGGGCGGAGGAGCGGATCCGGGACGA 240
241 GCTGCTCAGCGGTGGTGGGAGGCGCGGAGGCGGGCGGGCGGAGGAGCGGGCG 300
241 GCTGCTCAGCGGTGGTGGGAGGCGGGCGGGCGGGCGGGCGGGCGGAGGAGCG 300
301 GGGCGGGGAGCGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
301 GGGCGGGGAGCGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
361 TACTGTACATCGCGTTCATCGCATGGCCATCGGAGTTCGGCGGCGGGCGGCTT 420
361 TACTGTACATCGCGTTCATCGCATGGCCATCGGAGTTCGGCGGCGGGCGGCTT 420
421 CTGGCGAGATCAACGAGTACCTCATGGCAAGTTCCCTTTTCGGCGGAGTACAG 480
421 CTGGCGAGATCAACGAGTACCTCATGGCAAGTTCCCTTTTCGGCGGAGTACAG 480
481 GGCTGGCGCAACTCCGTGGCGCCCAACCTTTTCGCTCAACGAGTTCGTTCAAG 540
481 GGCTGGCGCAACTCCGTGGCGCCCAACCTTTTCGCTCAACGAGTTCGTTCAAG 540
541 CGCGACCTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
541 CGCGACCTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
601 TACACCTTCGCGGAGCGGGGCTTTCCGCGCGCGCGCGGCGGCGGCGGCGG 660
601 TACACCTTCGCGGAGCGGGGCTTTCCGCGCGCGCGCGGCGGCGGCGGCGG 660
661 GTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
661 GTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
721 GCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 780
721 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
781 AGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
781 AGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
841 CGCAGCGGTGCGCTCAGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 900
841 CGCAGCGGTGCGCTCAGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 900
901 TGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
901 TGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
961 CCGCTCTGCGGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
1021 CCACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080
1021 CCACCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1080
1081 CTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
1081 CTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
1141 CTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
1141 CTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
1201 CTAGCTTGA 1209
1201 CTAGCTTGA 1209
```

```
RESULT 2
US-10-274-177-9
; Sequence 9, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1161)..(1161)
; OTHER INFORMATION: n=a, c, g, or t
US-10-274-177-9

Query Match      100.0%; Score 1209; DB 12; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTTGGAGGTTCGTCCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
DB 1 ATGAAGTTGGAGGTTCGTCCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60

QY 61 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 61 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120

QY 121 GGTCTCAGATGGGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 121 GGTCTCAGATGGGAGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GCGGACGCGGAGACAGAGTTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 181 GCGGACGCGGAGACAGAGTTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 241 GCTGTGTCAGCGGTTCGCGGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCG 300
DB 241 GCTGTGTCAGCGGTTCGCGGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCG 300

QY 301 GCGCGGGGAGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 360
DB 301 GCGCGGGGAGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 360

QY 361 TACTGTATCATCGCGCTCATCGCCATGGCCATCGCGGCGGCGGCGGCGGCGG 420
DB 361 TACTGTATCATCGCGCTCATCGCCATGGCCATCGCGGCGGCGGCGGCGGCGG 420

QY 421 CTGGCGGAGATCAACGAGTACCTCATGGGCGAAGTTCCCTTTTCGGCGGCGGCGG 480
DB 421 CTGGCGGAGATCAACGAGTACCTCATGGGCGAAGTTCCCTTTTCGGCGGCGGCGG 480

QY 481 GGTGCGGCGCAACTCCGTTCGCGCGCAACCTTTTCGCTCAACGAGTTCGTTCA 540
DB 481 GGTGCGGCGCAACTCCGTTCGCGCGCAACCTTTTCGCTCAACGAGTTCGTTCA 540

QY 541 GCGGACCCCTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 541 GCGGACCCCTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

QY 601 TACACCTTCGCGGAGCGGGGCTTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCG 660
DB 601 TACACCTTCGCGGAGCGGGGCTTCGCGCGCGCGCGGCGGCGGCGGCGGCGGCG 660

QY 661 GTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
```

661 GTCCCGCGCCGCGGCTGCGGCGCGGAGGCGCCCGGCGCTCCCGCGCGCGCGCGCC 720
721 GCGCCCGCGCCCGCGGCTCGCCCGCATCGCTCGCCCGCGCGCGCGCGCGCGCGCG 780
721 GCGCCCGCGCCCGCGGCTCGCCCGCATCGCTCGCCCGCGCGCGCGCGCGCGCGCG 780
781 AGCCCGCGGCGAGTCTCCAGCTCTCTGCGCATCGACGATCTGCGCGAGCCCTTC 840
781 AGCCCGCGGCGAGTCTCCAGCTCTCTGCGCATCGACGATCTGCGCGAGCCCTTC 840
841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCC 900
901 TGCCCGCGCTGCGCGGCTTCGCGGCGCTCTCCCGCGCGCGCGCGCGCGCGCTGCTG 960
901 TGCCCGCGCTGCGCGGCTTCGCGGCGCTCTCCCGCGCGCGCGCGCGCGCGCTGCTG 960
961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCTGCGCGCGCGCGCGCGAGGTG 1020
961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCTGCGCGCGCGCGCGCGAGGTG 1020
1021 CCACCGACCGCGCGCGCGCTCTGCTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCG 1080
1021 CCACCGACCGCGCGCGCGCTCTGCTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCG 1080
1081 CTCGAGCG 1140
1081 CTCGAGCG 1140
1141 CTCGAGCG 1200
1141 CTCGAGCG 1200
1201 CTAGCTTGA 1209
1201 CTAGCTTGA 1209

RESULT 3
US-10-087-080-31
; Sequence 31. Application US/10087080
; Publication No. US20030235820A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Markowitz, Sanford David
; APPLICANT: Eos Biotechnology, Inc.
; APPLICANT: Case Western Reserve University
; TITLE OF INVENTION: No. US20030235820A1 Methods of Diagnosis of Metastatic Colorectal Cancer
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer
; FILE REFERENCE: 018501-000840US
; CURRENT APPLICATION NUMBER: US/10/087,080
; CURRENT FILING DATE: 2002-10-25
; PRIOR FILING DATE: 2002-10-25
; PRIOR FILING DATE: 2001-02-27
; PRIOR FILING DATE: 2001-02-27
; PRIOR FILING DATE: 2001-04-02
; PRIOR FILING DATE: 2001-04-02
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1161)
; OTHER INFORMATION: n = g, a, c or t

US-10-087-080-31
Query Match 100.0%; Score 1209; DB 15; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAGTTGGAGGTGTTCTGCTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 1 ATGAAGTTGGAGGTGTTCTGCTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Qy 61 GAGGCG 120
Db 61 GAGGCG 120
Qy 121 GGTCTCAGATGGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 121 GGTCTCAGATGGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy 181 GCGCAGCG 240
Db 181 GCGCAGCG 240
Qy 241 GCTGCTGAGCGGTGCTGCG 300
Db 241 GCTGCTGAGCGGTGCTGCG 300
Qy 301 GCG 360
Db 301 GCG 360
Qy 361 TACTGTATCATCGGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 361 TACTGTATCATCGGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Qy 421 CTGGCGGAGATCAACGAGTACTCTCATGGGCAAGTTCCCTTTTTCGCGCGCGCGCGCG 480
Db 421 CTGGCGGAGATCAACGAGTACTCTCATGGGCAAGTTCCCTTTTTCGCGCGCGCGCGCG 480
Qy 481 GGTGCG 540
Db 481 GGTGCG 540
Qy 541 GCGGACCGCTCG 600
Db 541 GCGGACCGCTCG 600
Qy 601 TACACCTTCG 660
Db 601 TACACCTTCG 660
Qy 661 GTCCCG 720
Db 661 GTCCCG 720
Qy 721 GCG 780
Db 721 GCG 780
Qy 781 AGCCCGCGGCGAGTCTTCAGCTCTCTGCGCATCGACGATCTTCGCGAGCGCGCTTC 840
Db 781 AGCCCGCGGCGAGTCTTCAGCTCTCTGCGCATCGACGATCTTCGCGAGCGCGCTTC 840
Qy 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCG 900
Db 841 CGCAGCGCTCGCTCAGGAGACAGCGCCCGCGGACGACGCTTCAGTGGGCGCGCGCGCG 900
Qy 901 TGCCCGCGCTGCGCGGCTTCGCGGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCTGCTG 960
Db 901 TGCCCGCGCTGCGCGGCTTCGCGGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCTGCTG 960
Qy 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCGCTGCGCGCGCGCGCGCGAGGTG 1020
Db 961 CCGCTCTGCGGTACGCGCGCGGCGAGCGCGCGCGCGCTGCGCGCGCGCGCGCGAGGTG 1020

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: SWISSPROT HIT: Q60688, EVALUE 1.00e-58
OTHER INFORMATION: EST HUMAN HIT: AW236501.1, EVALUE 3.00e-76
OTHER INFORMATION: NT HIT: U13223.1, EVALUE 0.00e+00
US-10-029-386-23708

Query Match 2.5%; Score 30; DB 14; Length 312;
Best Local Similarity 100.0%; Pred. No. 0.00037;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
Db 11 AAGCCCCCTACTCGTACATCGCGCTCATC 40

RESULT 7

US-10-027-632-138699
Sequence 138699, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 138699
LENGTH: 425
TYPE: DNA
ORGANISM: Human
US-10-027-632-138699

Query Match 2.5%; Score 30; DB 15; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

RESULT 8

US-10-027-632-138700
Sequence 138700, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 138700
LENGTH: 425
TYPE: DNA
ORGANISM: Human
US-10-027-632-138700

Query Match 2.5%; Score 30; DB 15; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

RESULT 9

US-10-027-632-138701
Sequence 138701, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 138701
LENGTH: 425
TYPE: DNA
ORGANISM: Human
US-10-027-632-138701

Query Match 2.5%; Score 30; DB 15; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
Db 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

RESULT 10

US-10-027-632-138703
Sequence 138703, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 138703
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-138703

Query Match          2.5%; Score 30; DB 15; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
DB 91 AAGCCCCCTACTCGTACATCGCGCTCATC 120

RESULT 11
US-10-029-386-10008
; Sequence 10008, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10008
; LENGTH: 599
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
; OTHER INFORMATION: NT HIT: U13223.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q60688, EVALUE 2.00e-63
; OTHER INFORMATION: EST_HUMAN HIT: AW236501.1, EVALUE 5.00e-76
US-10-029-386-10008

Query Match          2.5%; Score 30; DB 14; Length 599;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 AAGCCCCCTACTCGTACATCGCGCTCATC 381
```

```
DB 248 AAGCCCCCTACTCGTACATCGCGCTCATC 277

RESULT 12
US-09-867-701-3098
; Sequence 3098, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Agate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3098
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-701-3098

Query Match          2.1%; Score 25; DB 9; Length 441;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1181 CGTACCGGTGGAGACGCTCTTAGC 1205
DB 13 CGTACCGGTGGAGACGCTCTTAGC 37

RESULT 13
US-10-029-386-20506/c
; Sequence 20506, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20506
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008403.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: g113650682, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q03391, EVALUE 1.00e-61
; OTHER INFORMATION: EST_HUMAN HIT: B1912691.1, EVALUE 1.20e-01
US-10-029-386-20506

Query Match          2.1%; Score 25; DB 14; Length 1335;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 GGGGCCAGCGCGCGCGCGCGG 309
DB 796 GGGGCCAGCGCGCGCGCGCGG 772
```

RESULT 14
US-09-960-706-947
; Sequence 947, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Munger, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; TITLE OF INVENTION: Gene Expression Profiles
; FILE REFERENCE: 44921-5029-01US
; CURRENT APPLICATION NUMBER: US/09/960,706
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 60/223,323
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: 09/873,319
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 947
; LENGTH: 2187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 UI3220
US-09-960-706-947

Query Match 2.1%; Score 25; DB 10; Length 2187;
Best Local Similarity 100.0%; Pred. No. 0.057; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

QY 358 CCTACTCGTACATCGCGTCTATCG 382
|||||
Db 418 CCTACTCGTACATCGCGTCTATCG 442
|||||

RESULT 15
US-09-945-901-57
; Sequence 57, Application US/09945901
; Patent No. US20020161215A1
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lortie P.
; Ellis, Steven B.
; Liaw, Chen W.
; Lu, Chin-Chun
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; City: La Jolla
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/945,901
; FILING DATE: 24-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/940,035
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/052,449
; FILING DATE: 20-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-9393E
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999
TELEFAX: 619-238-0082
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4695 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 485..4495
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-09-945-901-57

Query Match 2.1%; Score 25; DB 9; Length 4695;
Best Local Similarity 100.0%; Pred. No. 0.049; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

QY 285 GGGGCCAGCGCGCGCGCGCGG 309
|||||
Db 3700 GGGGCCAGCGCGCGCGCGCGG 3724
|||||

Search completed: March 31, 2004, 06:38:09
Job time : 2479 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 02:16:13 ; Search time 2983 Seconds
(without alignments)

12103.043 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 1209

Sequence: 1 atgaagtggaggttgggt.....tggagacgtcctagcttga 1209

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hcto:*

9: gb_est1:*

10: gb_est2:*

11: gb_hct:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estum:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	31.1	904	13	BQ922461 AGENCOURT
2	146	12.1	982	13	BQ430047
3	131	10.8	246	29	CE688385 tigr-gss-
4	86	7.1	406	12	BMS10453 iJ41n03.y

5	77	6.4	398	9	AI586081
6	42	3.5	559	9	AI189632
7	38	3.1	232	12	BI402320
8	37	3.1	425	10	BF290883
9	37	3.1	514	10	BF282916
10	29	2.4	522	28	B56994
11	26	2.2	645	12	BI956079
12	26	2.2	750	13	BI719028
13	26	2.2	1069	13	BQ339574
14	25	2.1	156	29	CE072289
15	25	2.1	379	12	BM022612
16	25	2.1	401	10	BF564747
17	25	2.1	430	10	AW532942
18	25	2.1	441	9	AA292998
19	25	2.1	588	14	CF581582
20	25	2.1	927	13	BQ582829
21	25	2.1	1001	12	BI890470
22	25	2.1	1126	29	AG060944
23	25	2.1	1185	14	CF581581
24	24	2.0	272	9	AA048965
25	24	2.0	320	9	AI323408
26	24	2.0	347	10	BF523876
27	24	2.0	440	9	AI323053
28	24	2.0	442	28	BZ188256
29	24	2.0	646	14	CB576131
30	24	2.0	669	12	BI886947
31	24	2.0	808	14	CB960307
32	24	2.0	997	9	AL583668
33	24	2.0	1201	13	BX363285
34	23	1.9	129	14	CB245821
35	23	1.9	406	14	CA623788
36	23	1.9	518	10	BB637563
37	23	1.9	518	14	CA698771
38	23	1.9	533	12	BJ063635
39	23	1.9	592	12	BM156494
40	23	1.9	672	12	BG546422
41	23	1.9	698	29	AG060332
42	23	1.9	717	12	BG489468
43	23	1.9	774	14	CK148319
44	23	1.9	790	12	BG716647
45	23	1.9	898	12	BG752658

ALIGNMENTS

RESULT 1
BQ922461
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ922461 904 bp mRNA linear EST 20-AUG-2002
AGENCOURT 8963457 Lupeki sciatic_nerve Homo sapiens cDNA clone
IMAGE:6200329 5', mRNA sequence.
BQ922461
BQ922461.1 GI:22337492
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 904)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: Llam13615 row: a column: 02
High quality sequence stop: 523.

FEATURES
 source
 Location/Qualifiers
 1. 904
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6200329"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski sciatic nerve"
 /note="Vector: PCMV-SPORT6 (Life Technologies); Site 1:
 NorI; Site 2: SalI; cDNA made by oligo-dT priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCATCAGCGTCCG-3' and
 5'-GACATGTCATGTCGAGCGCCGCTT(15)-3'. Size selected >
 1 kb for average insert length 1.87 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life
 Technologies."

ORIGIN
 Query Match 31.1%; Score 376; DB 13; Length 904;
 Best Local Similarity 99.6%; Pred. No. 1.9e-132;
 Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 529 GTCAAGTGTGTCGCGACCCCTCCGCGCCCTGGGGGAGGACAACTACTGATGCTCAAC 588
 Db 42 GTCAAGTGTGTCGCGACCCCTCCGCGCCCTGGGGGAGGACAACTACTGATGCTCAAC 101
 Qy 589 CCCAACAGCGAGTACACCTTCGCGGAGCGGGGTCTTCGCGCGCGCCGCGCAAGCGCCCTCAGC 648
 Db 102 CCCAACAGCGAGTACACCTTCGCGGAGCGGGGTCTTCGCGCGCGCCGCGCAAGCGCCCTCAGC 161
 Qy 649 CACCGCGCGCGGTCCCGCGCCCGCGGTGCGGCGCGCGAGAGCGCCCGCGCTCCCGGCC 708
 Db 162 CACCGCGCGCGGTCCCGCGCCCGCGGTGCGGCGCGCGAGAGCGCCCGCGCTCCCGGCC 221
 Qy 709 GCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCATGCGTCCGCGCGCGCGCGCG 768
 Db 222 GCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCGCGCATGCGTCCGCGCGCGCGCGCG 281
 Qy 769 GAGAGCGCGCGAGCGCGCGCGCGAGTTCTCCAGCTCTTGGCCATCGACAGATCCTG 828
 Db 282 GAGAGCGCGCGAGCGCGCGCGCGAGTTCTCCAGCTCTTGGCCATCGACAGATCCTG 341
 Qy 829 CGCAAGCGCTTCCGACGCGTCCGCTCAGGACACGCGCGCGGAGCGCTTCAGTGG 888
 Db 342 CGCAAGCGCTTCCGACGCGGCGCTCAGGACACGCGCGCGGAGCGCTTCAGTGG 401
 Qy 889 GCGCGCGCGCGCTCCCGCGCGCTCCCGCGTTCGCGCGCGCTCTCTCCCGCGCGCGCG 948
 Db 402 GCGCGCGCGCGCTCCCGCGCGCTCCCGCGTTCGCGCGCGCTCTCTCCCGCGCGCGCG 461
 Qy 949 AGGCGCTGCTGCGCTCTGCGCTACGCGCGCGCGGAGCGCGCGCGCTGGCGGC 1006
 Db 462 AGGCGCTGCTGCGCTCTGCGCTACGCGCGCGCGGAGCGCGCGCGCTGGCGGC 519

RESULT 2
 BQ430047 982 bp mRNA linear EST 24-MAY-2002
 LOCUS AGNCOURT_7916258 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6010412
 DEFINITION 5', mRNA sequence.
 BQ430047
 VERSION BQ430047.1 GI:21169123
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 982)

REFERENCE
 1 (bases 1 to 982)

AUTHORS
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue procurement: DCTD/BTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 DNA distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM13198 row: g column: 21
 High quality sequence stop: 491.
 Location/Qualifiers
 1. 982
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6010412"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_68"
 /note="Organ: lung; Vector: PCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

ORIGIN
 Query Match 12.1%; Score 146; DB 13; Length 982;
 Best Local Similarity 100.0%; Pred. No. 5.4e-45;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATGAAGTTGGAGTGTTCGTCCTCGCGCGCCACAGCGGAGCAGTGCCTG 60
 Db 265 ATGAAGTTGGAGTGTTCGTCCTCGCGCGCCACAGCGGAGCAGTGCCTG 324
 Qy 61 GAGGCGCGCGCGCGCGCGCGCGTCCCGCTGTCGCGCGGAGAGACTCCCTG 120
 Db 325 GAGGCGCGCGCGCGCGCGCGTCCCGCTGTCGCGCGGAGAGACTCCCTG 384
 Qy 121 GCTCAGATGGGAGTGGCGGCGCA 146
 Db 385 GCTCAGATGGGAGTGGCGGCGCA 410

RESULT 3
 CE688385 246 bp DNA linear GSS 29-SEP-2003
 LOCUS tigr-gss-dog-17000368532929 Dog Library Canis familiaris genomic,
 DEFINITION Genomic survey sequence.
 CE688385
 VERSION CE688385.1 GI:37007455
 KEYWORDS GSS.
 SOURCE Canis familiaris (dog)
 ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 246)
 Kirness,E.F., Baina,V., Halpern,A.L., Levy,S., Remington,K.,
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
 Venter,J.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 22875432
 14512627
 Contact: Kirness EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirnes@tigr.org

```

Class: shotgun.
Location/Qualifiers
1. 246
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 10.8%; Score 131; DB 29; Length 246;
Best Local Similarity 100.0%; Pred. NO. 2.5e-39;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 CTCACGACGCTTCGTCAGGTCGTCGCGACCCCTCGCGGCCCTCGGGCAGGACAC 573
DB 1 CTCACGACGCTTCGTCAGGTCGTCGCGACCCCTCGCGGCCCTCGGGCAGGACAC 60
QY 574 TACTGGATGTCACCCCAACAGCAGTACACCTTCGCCGACGGGTCTTCGCCGCCGC 633
DB 61 TACTGGATGTCACCCCAACAGCAGTACACCTTCGCCGACGGGTCTTCGCCGCCGC 120
QY 634 CGCAGCGCCT 644
DB 121 CGCAGCGCCT 131

RESULT 4
BM510453 406 bp mRNA linear EST 15-FEB-2002
LOCUS 1j41h03.y1 Human insulinoma Homo sapiens cdna clone IMAGE:5633525
DEFINITION 5, similar to TR:070220 070220 FORK HEAD TRANSCRIPTION FACTOR. 1,
mRNA sequence.
ACCESSION BM510453
VERSION BM510453.1 GI:186681596
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 406)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemihka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvili,R.,
Williams,T., Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Other ESTs: i41h03.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioph.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seg primer: -40RP from Gibco
High quality sequence stop: 398.
Location/Qualifiers
1. 406
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cGene="IMAGE:5633525"
/tissue_type="insulinoma"

FEATURES
source
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; site 2: EcoRI; Constructed with lambda ZapII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
9127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."

ORIGIN
Query Match 7.1%; Score 86; DB 12; Length 406;
Best Local Similarity 100.0%; Pred. No. 3.4e-22;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 CCCGCGCAAGCCTCCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1127
DB 48 CCCGCGCAAGCCTCCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 107
QY 1128 GCTGCCCGCAGCCCTGCAGCGCGCT 1153
DB 108 GCTGCCCGCAGCCCTGCAGCGCGCT 133

RESULT 5
AI586081 398 bp mRNA linear EST 06-APR-1999
LOCUS v092c05.x1 Stratagene mouse skin (#937313) Mus musculus cdna clone
DEFINITION IMAGE:1209608 3', similar to gb:U36760 Mus musculus brain factor-1
(MOUSE);, mRNA sequence.
ACCESSION AI586081
VERSION AI586081.1 GI:4571978
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 398)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:645952
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 387.
Location/Qualifiers
1. 398
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1209608"
/sex="females"
/tissue_type="whole skin"
/dev_stages="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse skin (#937313)"

```


/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

ORIGIN

Query Match 6.4%; Score 77; DB 9; Length 398;
Best Local Similarity 99.2%; Pred. No. 8.8e-19;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTCGTCAGGTCCTCGGAGCCCTCGCGGCCCTGGGCGAGGACAACTACTGATGCTC 585
|||||
Db 44 TTGTCAGAGTGCTCGGACCCCTCGCGGCCCTGGGCGAGGACAACTACTGATGCTC 103
|||||

QY 586 AACCCCAACAGCAGTACACCTTCGCCGACGGGGTCTTCGCCCGCCGCGCAAGCGCTC 645
|||||
Db 104 AACCCCAACAGCAGTACACCTTCGCCGACGGGGTCTTCGCCCGCCGCGCAAGCGCTC 163
|||||

QY 646 AGCCACCG 653
|||||
Db 164 AGCCACCG 171
|||||

RESULT 6

Al169632/c
LOCUS
DEFINITION EST215519 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone
RKBV46 3' end, mRNA sequence.
ACCESSION Al169632
VERSION Al169632
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.

REFERENCE
AUTHORS Lee, N.H., Glöck, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES

Location/Qualifiers
1..559
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2027517"
/db_xref="taxon:10118"
/clone="RKIBV46"
/note="Organ: kidney; Vector: pT73Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN

Query Match 3.5%; Score 42; DB 9; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TACACTTCGCGAGCGGGTCTTCGCCCGCCGCGCGCAAGCGC 642
|||||
Db 548 TACACTTCGCGAGCGGGTCTTCGCCCGCCGCGCGCAAGCGC 507
|||||

RESULT 7

BI402320
LOCUS
DEFINITION MI-P-CP0-nwd-f-07-0-UI-s1 MI-P-CP0 Sus scrofa cDNA clone
ACCESSION BI402320
VERSION BI402320
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
TITLE 1 (bases 1 to 232)
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE discovery
PUBMED Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
8889548

Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Towa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab, University of Iowa EST sequencing: M.B. Soares Lab,
University of Iowa Clone Distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.

FEATURES

Location/Qualifiers
1..232
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="MI-P-CP0-nwd-f-07-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="MI-P-CP0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI. The MI-P-CP0
library is derived from uterus. For a detailed description
of the library from which this clone was derived, please
visit our web site at http://pigest.genome.iastate.edu/.
The procedure used to create this library has been
previously described (Bonaldo, Lennon and Soares, Genome
Research 6:791-806, 1996)
TAG_SEQ=None found"

ORIGIN

Query Match 3.1%; Score 38; DB 12; Length 232;
Best Local Similarity 100.0%; Pred. No. 0.00058;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 TCGCTCAACGACTGCTTGTCAAGTGCTGCGCGACCC 548
|||||
Db 105 TCGCTCAACGACTGCTTGTCAAGTGCTGCGCGACCC 142
|||||

RESULT 8

BF290883/c
LOCUS
DEFINITION BF290883 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
ACCESSION BF290883
VERSION BF290883
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE
AUTHORS      1 (bases 1 to 425)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Perlea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE        Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL      Unpublished (2000)
COMMENT      Other ESTs: EST353025
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
Location/Qualifiers
1..425
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RG11F65"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RKL, RLI, RPL, RLU, REM, RMU,
RSP, RHE, RPC, RPN"
ORIGIN
Query Match      3.1%; Score 37; DB 10; Length 425;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GCAAGTTCCTCCAGCTCTTCGCCATCGACAGCATCCT 827
Db 343 GCAAGTTCCTCCAGCTCTTCGCCATCGACAGCATCCT 307

RESULT 9
BF282916/c
LOCUS      BF282916      514 bp      mRNA      linear      EST 28-NOV-2000
DEFINITION Rattus norvegicus cDNA clone RGIDU92 3' sequence, mRNA sequence.
VERSION    BF282916.1 GI:11213986
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE  1 (bases 1 to 514)
AUTHORS    Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Perlea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE      Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL    Unpublished (2000)
COMMENT    Other ESTs: EST348098
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
Location/Qualifiers
1..514
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="RGIDU92"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"

```

```

/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/notes="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RKL, RLI, RPL, RLU, REM, RMU,
RSP, RHE, RPC, RPN"
ORIGIN
Query Match      3.1%; Score 37; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 791 GCAAGTTCCTCCAGCTCTTCGCCATCGACAGCATCCT 827
Db 356 GCAAGTTCCTCCAGCTCTTCGCCATCGACAGCATCCT 320

RESULT 10
B56994
LOCUS      B56994      522 bp      DNA      linear      GSS 20-JUN-1998
DEFINITION CIT-HSP-2008G17.TR CIT-HSP Homo sapiens genomic clone 2008G17,
genomic survey sequence.
ACCESSION  B56994
VERSION    B56994.1 GI:2611328
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 522)
AUTHORS    Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)
Other GSSs: CIT-HSP-2008G17.TFB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: madams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
Location/Qualifiers
1..522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7040419"
/db_xref="taxon:9606"
/clone="2008G17"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-Hsp"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
ORIGIN
Query Match      2.4%; Score 29; DB 28; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 AGCCCCCTACTCGTACATCGCGCTCATC 381
Db 488 AGCCCCCTACTCGTACATCGCGCTCATC 516

RESULT 11
B1956079/c

```

LOCUS B1956079 645 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEM0025J14f Hordeum vulgare green seedling EST library
clone HVCDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
VERSION B1956079
KEYWORDS B1956079.1 GI:16302947
SOURCE EST.
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 645)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 202
Seq primer: AATTAACCTCCTCACTAAAGGG
High quality sequence start: 32
High quality sequence stop: 409.
FEATURES
source
1. 645
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clones="HVSMEM0025J14f"
/tissue_type="green seedling leaf"
/lab_host="TUC121"
/clone_lib="Hordeum vulgare green seedling EST library"
HVCDNA0014 (Blumeria infected)"
/note="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the TJ Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/ggp/bgn/31/cover.html>)"
ORIGIN
Query Match 2.2%; Score 26; DB 12; Length 645;

Best Local Similarity 100.0%; Pred. No. 22;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 699 CTTCCCGCGCGCGCGCGCGCGCGCGC 724
|||||
DB 598 CTTCCCGCGCGCGCGCGCGCGCGCGC 573
|||||
RESULT 12
LOCUS BX719028 750 bp mRNA linear EST 18-NOV-2003
DEFINITION BX719028 XGC-tadpole Silurana tropicalis cDNA clone TTPA063f13 5',
mRNA sequence.
ACCESSION BX719028
VERSION BX719028
KEYWORDS BX719028.1 GI:38391769
SOURCE EST.
ORGANISM Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 750)
AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TTPA063f13.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
cDNA was oligo dt primed from sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.
FEATURES
source
1. 750
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TTPA063f13"
/dev_stage="tadpole (stage 35-40)"
/lab_host="E. coli DH10B"
/clone_lib="XGC-tadpole"
/note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from tadpole
embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
with EcoRI at the 5' end and NotI at the 3' end"
ORIGIN
Query Match 2.2%; Score 26; DB 13; Length 750;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 352 AAGCCCCCTACTCGTACATCGCGCT 377
|||||
DB 335 AAGCCCCCTACTCGTACATCGCGCT 360
|||||
RESULT 13
LOCUS BQ939574/c 1069 bp mRNA linear EST 21-AUG-2002
DEFINITION BQ939574 AGENCOURT_8931588 NTH_MGC_40 Homo sapiens cDNA clone IMAGE:6484117
5' mRNA sequence.
ACCESSION BQ939574
VERSION BQ939574.1 GI:22355052
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1069)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2669 row: m column: 14
High quality sequence stop: 228.
Location/Qualifiers
1..1069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:8484117"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Query Match 2.2%; Score 26; DB 13; Length 1069;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 707 CGCCCCCGCGCGCGCGCGCGCGCGCG 732
Db 880 CGCCCCCGCGCGCGCGCGCGCGCGCG 855

RESULT 14
CE072289/c
LOCUS
DEFINITION
tigr-gss-dog-17000323549470 Dog Library Canis familiaris genomic,
genomic survey sequence.
CE072289
VERSION
CE072289.1 GI:351335052
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 156)
Kirkness, E.F., Baina, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
14512627
PUBLISHED
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

```

```

FEATURES
source
Location/Qualifiers
1..156
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 2.1%; Score 25; DB 29; Length 156;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 CCTACTCGTACATCGGCTCATCG 382
Db 140 CCTACTCGTACATCGGCTCATCG 116

RESULT 15
BM022612/c
LOCUS
DEFINITION
BM022612 379 bp mRNA linear EST 12-MAR-2002
1362H09.x1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus
musculus cDNA clone IMAGE:5668072 3' similar to SW:PRE1_MOUSE
Q61080 FORKHEAD-RELATED TRANSCRIPTION FACTOR 1, mRNA sequence.
BM022612
ACCESSION
BM022612.1 GI:16536968
VERSION
EST.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 379)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lamishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hiller, J., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R.,
Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
MGI:1954398 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
High quality sequence stop: 374.
Location/Qualifiers
1..379
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5668072"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMS1"
/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5

```

pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using Superscript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:731-806, 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

ORIGIN

Query Match	2.1%	Score 25;	DB 12;	Length 379;
Best Local Similarity	100.0%;	Pred. NO. 52;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	358	CCCTACTCGTACATCGCGTCTATCG	382	
DB	266	CCCTACTCGTACATCGCGTCTATCG	242	

Search completed: March 31, 2004, 04:42:53
Job time : 2986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 29, 2004, 11:52:41 ; Search time 64.5 Seconds
(without alignment)
10592.246 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189

Sequence: 1 atgaagtggaggtgttcgt.....tggagagcctcctagcttga 1209

Scoring table:
XGAPUP 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/UGPTO.spool -P/US10087080/runat 29032004 114525 6520/app_query.fasta_1.1351
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080 @CGN 1.1.81/runat 29032004 114525 6520 -NCPU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databases : A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003s.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2123	97.0	402	5 ABP54691	Abp54691 Metastati
2	1314.5	60.1	439	7 ADD48790	Add48790 Rat Prote
3	1314.5	60.1	439	7 ADD47621	Add47621 Rat Prote
4	517	23.6	465	7 ADB75312	Adb75312 Prostate
5	467	21.3	478	6 ABR63544	Abr63544 Danio rer
6	467	21.3	478	7 ADD48650	Add48650 Human Pro
7	467	21.3	478	7 ADD46993	Add46993 Human Pro
8	467	21.3	478	7 ADD46997	Add46997 Human Pro
9	466	21.3	464	6 ABR63545	Abr63545 Danio rer
10	445	20.3	396	6 ABR63546	Abr63546 Danio rer

11	436.5	19.9	409	6 ABU56559	Abu56559 Lung canc
12	431.5	19.7	553	2 AAY43260	Aay43260 Human FKX
13	431.5	19.7	553	2 AAY55721	Aay55721 Human for
14	429.5	19.6	553	2 AAY41277	Aay41277 Human FRE
15	415.5	19.0	376	5 ABJ01138	Abj01138 Ovary cel
16	402.5	18.4	478	4 ABR68496	Abr68496 Novel hum
17	402	18.4	371	6 ABR63547	Abr63547 Danio rer
18	394.5	18.0	330	4 AAB61327	Aab61327 Human tra
19	393	18.0	465	4 ABG03158	Abg03158 Novel hum
20	387.5	17.7	501	2 AAY01097	Aay01097 FREAC11 p
21	387	17.7	473	5 AAU74806	Aau74806 Human hep
22	387	17.7	473	5 AAU74807	Aau74807 Human hep
23	387	17.7	473	6 ABR47488	Abr47488 Breast ca
24	384.5	17.6	501	4 AAB82851	Aab82851 Human win
25	384.5	17.6	501	5 AAU79816	Aau79816 Human tra
26	384.5	17.6	501	6 ADA50747	Ada50747 Human FOX
27	384	17.5	494	5 AAU79818	Aau79818 Mouse mes
28	373	17.0	19938	6 ABP76680	Abp76680 Streptomy
29	366.5	16.3	19938	6 ABP76680	Abp76680 Streptomy
30	366	16.7	371	6 ABR63543	Abr63543 Danio rer
31	366	16.7	428	4 ABR63539	Abr63539 Craniofac
32	365	16.7	428	4 ABR63539	Abg08497 Novel hum
33	360	16.4	19938	6 ABP76681	Abp76681 Streptomy
34	354.5	16.2	456	4 ABB60077	Abb60077 Drosophi
35	347.5	15.9	19938	6 ABP76682	Abp76682 Streptomy
36	345	15.8	480	2 AAR44551	Aar44551 Brain fac
37	342.5	15.6	19938	6 ABP76679	Abp76679 Streptomy
38	341	15.6	481	5 ABB57076	Abp72344 Mouse isc
39	338.5	15.5	477	6 ABR72344	Abp72344 Brain fac
40	338	15.4	509	5 ABG95654	Abg95654 Human nuc
41	336	15.0	19938	6 ABP76681	Abp76681 Streptomy
42	333.5	15.2	182	5 ABP43874	Abp43874 Transcrip
43	332.5	15.2	19938	6 ABP76678	Abp76678 Streptomy
44	329	15.0	544	2 AAU90247	Aau90247 Human PAS
45	326.5	14.9	451	4 ABB59467	Abb59467 Drosophi

ALIGNMENTS

RESULT 1
ABP54691
ID ABP54691 standard; protein; 402 AA.
XX
AC ABP54691;
XX
DT 30-DEC-2002 (first entry)
XX
DE Metastatic colorectal cancer-associated polypeptide.
XX
KW Colorectal cancer; metastasis; differential expression; cytostatic;
XX
KW diagnosis; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 387
FT /note= "encoded by CGN"
XX
PN WO200268677-A2.
XX
PD 06-SEP-2002.
XX
PF 27-FEB-2002; 2002WO-US006001.
XX
PR 27-FEB-2001; 2001US-0272206P.
PR 02-APR-2001; 2001US-0281149P.
PR 17-APR-2001; 2001US-0284555P.
XX
PA (E0SB-) EOS BIOTECHNOLOGY INC.
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Mack DH, Markowitz SD;
XX

DR GENBANK; Q63244.
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX
PS Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 439 AA;

Alignment Scores:
Pred. No.: 1,096-74 Length: 439
Score: 1314.50 Matches: 278
Percent Similarity: 74.30% Conservative: 14
Best Local Similarity: 70.74% Mismatches: 86
Query Match: 60.05% Indels: 15
DB: 7 Gaps: 6

US-10-087-080-31 (1-1209) x ADD48790 (1-439)

QY 49 GGCAGTACCTGGAGGCGCGGGCGGAGCGAGCGCGCTCCCGCTGTCGGCGCGGGA 108
DB 2 GlySerAspLeuGluGlyAlaGlySerSerAspValProSerProLeuSerAlaAlaGly 21
QY 109 GACGACTCCCTGGCTCAGATGGGACTGCGCGCC---AAGCCGTCCGCGGCGCGCG 165
DB 22 AspAspSerLeuGlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlySerGly 41
QY 166 GCCAGATACCGAG---GGCAGCGGCAACAGAGTGGGAGCGGCGCGCGCGCGGAG 222
DB 42 AlaGlyAspLeuGluGlyGlyGlyGluArgAsnSerSerGlyGlyAlaSerThrGln 61
QY 223 GAGCGGATCCCGGCGAGCAGCTCTCAGCGGTGTGGCGAGGGCGCG---GAGGCC 276
DB 62 Asp-----AspProGluValThrAspGlySerArgThrGlnAla 74
QY 277 GGGCGGCGGGGCGCAGGCGCGCGCGCGGCGGAGCGGAGGTGCACGAGCAAGCCA 336
DB 75 SerProValGlyProCysAlaGlySerValGlyGlyGlyGluAlaArgSerLysPro 94
QY 337 TATAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
DB 95 TyrThrArgGlyProCysAlaGlySerValGlyGlyGlyGlyGluAlaMetAlaAlaArg 114
QY 397 GACTCGGCGGCGGCGGCGGCTTGACGCTGGCGGAGATCAACAGTAGTACCTCATCGGAGTTTC 456

DB 115 AspSerAlaGlyGlyArgLeuThrLeuAlaGluIleAsnGluThrLeuMetGlyLysPhe 134
QY 457 CCCTTTTTCGCGCGAGCTACACGGGCTGGCGCAACTCCGTGGCCACCACTTCCTGCTC 516
DB 135 ProPhePheArgGlySerThrGlyTrpArgAsnSerValArgHisAsnLeuSerLeu 154
QY 517 AACGACTGCTTCGTCAGGTGCTCGCGACCTCCGGCCCTCGGGCGGCAAGGCACTAC 576
DB 155 AsnAspCysPheValLysValLeuArgAspProSerArgProTrpGlyLysAspAsnTyr 174
QY 577 TGGATGCTCAACCCCAACACGAGGAGTACACTTCGCGAGCGGGGTCTTCGCGCGCCCGC 636
DB 175 TrpMetLeuAsnProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgGly 194
QY 637 AACGGCTCAGCCAGCG 696
DB 195 LysArgLeuSerHisArgThrThrValSerAlaSerGlyTyrGlyGlySerProPro 214
QY 697 GGCTCCCGCTCG 756
DB 215 GlyProAlaGlyThrProGlnProAlaProThrAlaGlySerSerProIleAlaArgSer 234
QY 757 CCG 816
DB 235 ProAlaArgGlnGluGlySerSerProAlaSerLysPheSerSerSerPheAlaIle 254
QY 817 GACAGCATCTCGCGAAGCCCTTCGCGAGCGCTCAGGGACACAGCGCGCGCGCGCG 876
DB 255 AspSerIleLeuSerLysProPheArgSerArgArg-AspGlyThrArgLeuTrpGlyC 274
QY 877 ACCTTCAGTGGGCG 936
DB 274 sSerTyrProGlyAlaLeuLeuProAlaArgArgCysAlaProIleProArgSerPhePr 294
QY 937 GCGGCG 996
DB 294 oArgAspProAlaValProCysArgSerValLeuThrAlaArg-ArgAlaHisAla 314
QY 997 CTGGCG 1056
DB 314 laGlyValAlaArgGlyArgGlyAlaAla---ArgGlyAlaProValArgGlyAlaProL 333
QY 1057 CCG 1113
DB 333 euHisArgGlyProSerGlnAlaIleSerArgSerGlyAspArgArgGlyAlaProV 353
QY 1114 TACTGCG 1173
DB 353 alLeuProProThrAlaAlaHisGlyProAlaGlyArgGlyLeuArgSerGlySerA 373
QY 1174 CACTGTGTGTACCGGTGGAGCGCTCTAGCTTG 1208
DB 373 laProValLeuProGlyGlyAspAlaAlaSerLeu 384

RESULT 3
ADD47621
ID ADD47621 standard; protein; 439 AA.
XX
XX ADD47621;
XX
XX 29-JAN-2004 (first entry)
XX
XX Rat Protein Q63244, SEQ ID NO 13317.
XX
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
XX Rattus norvegicus.
XX
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX

RESULT 5
ABR63544
ID ABR63544 standard; protein; 478 AA.
XX
AC ABR63544;
XX
DT 18-SEP-2003 (first entry)
XX
DE Danio rerio foxd3 homologue human HPH2.
XX
DE Craniofacial malformation; variant foxd3; osteopathic; cardiant;
KW cytotatic; auditory; neuroprotective; cleft palate syndrome;
KW congenital heart disease; cancer; human; HPH2.
XX
OS Homo sapiens.
XX
PN WO2003048196-A2.
XX
PD 12-JUN-2003.
XX
PF 09-DEC-2002; 2002WO-EP013936.
XX
PR 07-DEC-2001; 2001EP-00129205.
XX
PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.
XX
PI Knapik E, Sachdev S;
XX
DR WPI; 2003-513737/48.
XX
PT New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing
PT and treating craniofacial malformations, such as cleft palate syndrome,
PT loss of middle ear ossicles, congenital heart disease, cancer and mental
PT retardation.
XX
PS Claim 34; Page 64-66; 70pp; English.
XX
CC The present invention provides the protein and coding sequences of Danio
CC rario foxd3. These can be used in the diagnosis and treatment of
CC craniofacial malformations, such as cleft palate syndrome, loss of middle
CC ear ossicles, congenital heart disease, Nishikawa-Kuroki syndrome, cancer
CC and mental retardation. The present sequence is a homologue of the
CC zebrafish foxd3 protein, and is human HPH2
XX
SQ Sequence 478 AA;

Alignment Scores:
Pred. No.: 2,09e-21 Length: 478
Score: 467.00 Matches: 161
Percent Similarity: 39.88% Conservative: 36
Best Local Similarity: 32.59% Mismatches: 164
Query Match: 21.33% Indels: 134
DB: 6 Gaps: 17

US-10-087-080-31 (1-1209) x ABR63544 (1-478)
QY 37 GGGGACAAGCAGGCGAGTGAACCTGGAGGCG-----GCGGGCGGCGAGCGAGCGCG 87
Db 5 GlyGlyGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24
QY 88 TCCCGCTGTCGGCGGCGGAGAGAGTCCCTGGGCTCAGATGGGACTGCGGCGGCAAG 147
Db 25 IleAspValValGlyGlyGlyAspGlyLeuGluGlyAspSerAspAlaGlyCys 44
QY 148 CCGTCCGCGGCGGCGGCGGCGAGATACGCGAGGCGAGCGGAACAGAGTG-----199
Db 45 AspSerProAlaGly-ProProGluLeuArgLeuAspGluAlaAspGluValProAla 64
QY 200 -----CGGA 204
Db 64 aAlaProHisGlyGlnProGlnProHisGlnGlnProLeuThrLeuProLysG 84

QY 205 GCGGCGCGGCGCGGCGA-----GGAGGCGAT-----CCTGGGCGAGCGTGTCT 246
Db 84 uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104
QY 247 GCAGCGGTGTGGCGGAGGCGC-----GGAGCGCGGCGCGGCGGCGCGCGGCG 300
Db 104 ySLysGlyGlyValGlyGlyGluGluGlyGlyAlaSerGlyGly-GlyProGlyAlaGly 123
QY 301 GCGCGGCGGAGCGGCGAGGTGTGACGAGCAAGCATATACGCGGCGGCGGCGGCGGCG 360
Db 124 SerGlySerAlaGlyLeuAlaProSerLysProLysAsnSerLeuValLysProPro 143
QY 361 TACTCGTACATCGGCTCATCGCATCGCATCGGACTCGGCGGCGGCGGCGGCGTTCAGC 420
Db 144 TyrSerTyrlleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLysLeuThr 163
QY 421 CTGCGGAGATCAACAGATACCTCATGGCAAGTTCCTTTTTCGCGGCGAGCTACACG 480
Db 164 LeuSerGlyIleCysGluPheIleSerAsnArgPheProTyTyArgGluLysPhePro 183
QY 481 GGCTGGCGCAACTCCGTGGCGCCACACTTTCCTCAACGACTGCTTCGTCAAGGTCTG 540
Db 184 AlaTrpGlnAsnSerIleA:GHisAsnLeuSerLeuAsnAspCysPheValLysIlePro 203
QY 541 CGGACCCCTCGCGGCGCTGGGCGAGCAAGCAACTACTGTGATCTCAACCCCAACAGCGAG 600
Db 204 ArgGluProGlyAsnPro---GlyLysGlyAsnTyTrpThrLeuAspProGlnSerGlu 222
QY 601 TACACCTTGGCGGAGCGGCTTCGCGCGCGCGGCGGAGCGCTCAGCCACCGCGCGCC- 659
Db 223 AspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLysArgHisGlnGln 242
QY 660 -----GGTCCCGCGCGCGCGCGCGCT 677
Db 243 GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTySerLeuAla 262
QY 678 GCGGCGGAGAGCGCGCGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
Db 263 AlaAlaAlaGlyAlaAlaGlyProTyArgPro-----TyrglyLeuHisProAla 280
QY 738 CTCGCCCGCGTGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAAGTT 797
Db 281 AlaAlaAlaGlyAlaTySerHisProAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 300
QY 798 CTCAGGTCTTCGCCCATTCGACGATCTCGGCAAGCCCTTCGCGAGCGGTG----- 851
Db 301 LeuGlnTyProTyAlaLeuProProValAlaProValLeuPro---ProAlaValPro 319
QY 852 -----CCTCAGGAGACAGCGCGCGCGCGCGCGCTTCAGTGGGCGCGCGCG- 896
Db 320 LeuLeuProSerGlyGluLeuGlyArgLysAlaAlaAlaAlaAlaAlaAlaAlaAla 339
QY 897 -----GCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT- 920
Db 340 GlyLeuGlnLeuGlnLeuAsnSerLeuGlyAlaAlaAlaAlaAlaAlaAlaAlaGly 359
QY 921 -----CCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 944
Db 360 AlaAlaGlyThrAlaSerLeuIleLysSerGluProSerAlaArgProSerPheSer 379
QY 945 CTGCAG-----GGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 989
Db 380 IleGluAsnIleIleGlyGlyGlyProAlaAlaPro-----GlyGlySerAla 395
QY 990 GCGCGGCTGGCG 1049
Db 396 ValGlyAlaGlyValAlaGlyGlyThrGlyGly-----406
QY 1050 ACCTCTCCCG 1106
Db 407 ---SerGlyGlyGlySerThrAlaGlnSerPheLeuArgProProGlyThrValGlnSer 425
QY 1107 -----GCA 1109

Db 426 AlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArgThrAlaThrIleAla 445
Qy 1110 CTTGACTG-----CCCTGGCGGTGGCCGAGC 1139
Db 446 ProIleLeuSerValProLeuSerGlyGlnPheLeuGlnProAlaAlaSerAlaAlaAla 465
Qy 1140 CTTGAGGCGGCTTAGTCGCGCTCTGGCCGCA 1175
Db 466 AlaAlaAlaAlaAlaGlnAlaLysTrpProAla 477

RESULT 6
ID ADD48650
XX ADD48650 standard; protein; 478 AA.
XX AC ADD48650;
XX XX
XX 29-JAN-2004 (first entry)
XX XX
XX Human Protein NP_036315, SEQ ID NO 14356.
XX KW Human; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; Chronic constriction injury; CCI;
XX KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX PN WO2003016475-A2.
XX XX
XX PD 27-FEB-2003.
XX XX
XX PF 14-AUG-2002; 2002MO-US025765.
XX XX
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX XX
XX PA (GEHO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX XX
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX XX
XX DR WPI; 2003-268312/26.
XX DR GENBANK; NP_036315.

XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the
XX specification, a method for identifying a compound useful in treating
XX pain and a pharmaceutical composition comprising the one or more
XX polypeptides or their antibodies. The polynucleotide or the compound that
XX modulates its activity is useful for preparing a medicament for treating
XX pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 478 AA;
Alignment Scores: 2.09e-21 Length: 478
Pred. No.: 467.00 Matches: 161
Score: 39.88% Conservative: 36
Percent Similarity: 32.59% Mismatches: 164
Best Local Similarity: 21.33% Indels: 134
Query Match: 7 Gaps: 17
DB: 17

US-10-087-080-31 (1-1209) x ADD48650 (1-478)
Qy 37 GGGGACAGCAGGGCAGTGTGAGTGGAGGCG-----GCGGGCGGACGCGGCGCG 87
Db 5 GlyGlyGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24
Qy 88 TCCCGCGTGTGCGGCGGAGAGCAGCTCCCTGGGTCTAGATGGGACTCGCGGCGCAAG 147
Db 25 IleAspValValGlyGlyGlyAspGlyLeuGluGlyLysAspSerAspAlaGlyCys 44
Qy 148 CCGTCCCGGCGGCGGCGGCGGAGATACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 199
Db 45 AspSerProAlaGly-ProProGluLeuArgLeuAspGluAlaAspGluValProAla 64
Qy 200 -----CGGA 204
Db 64 aAlaProHisGlyGlnProGlnProHisGlnGlnProLeuThrLeuProLysGly 84
Qy 205 GCGGGCGGCGGCGGCGG-----GGAGCGAT-----CCGGGCGAGCGTGTCT 246
Db 84 uAla-AlaGlyAlaGlyAlaGlyProGlyGlyAspValGlyAlaProGluAlaAspGlyC 104
Qy 247 GCAGCGTGTGTGGCGGAGGCGC-----GGAGCGCGGCGGCGGCGGCGGCGGCGG 300
Db 104 yLysGlyGlyValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 123
Qy 301 GCGCGGCGGAGCGGCGGAGGTGACGAGCAAGCATATACGCGGCGGCGGCGGCGGCGG 360
Db 124 SerGlySerAlaGlyGlyLeuAlaProSerLysProLysAsnSerLeuValLysPro 143
Qy 361 TACTGTGTACATCGGCTCATCGCATCGCATCGCATCGCATCGGCGGCGGCGGCGGCG 420
Db 144 TyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLysLeuThr 163
Qy 421 CTGGCGGAGATCAACGAGTACTCTCATGGCAAGTTCCTTTTTCGCGGCGGCGTACAG 480
Db 164 LeuSerGlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPhePro 183
Qy 481 GGCTGGCGCACTCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 184 AlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIlePro 203
Qy 541 CGGACCCCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 204 ArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGlu 222
Qy 601 TACACTTTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659
Db 223 AspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLysArgHisGlnGln 242
Qy 660 -----GCTCCCGCGGCGGCGGCGGCT 677
Db 243 GluHisLeuArgGluGlnThrAlaLeuMetMetGlnSerPheGlyAlaTyrSerLeuAla 262
Qy 678 GCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 737
Db 263 AlaAlaAlaGlyAlaAlaGlyProTyrGlyArgPro-----TyrGlyLeuHisProAla 280

QY	852	-----CCTCAGGACACGGCCCCCGGACGAGCTTCAGTGGGGCGCGC-----	896
Db	320	LeuLeuProSerGlyGluLeuGlyArgGlyAlaAlaAlaPheGlySerGlnLeuGlyPro	339
QY	897	-----GCCCTGCCCGCGCTCCCGCGCT-----	920
Db	340	GlyLeuGlnLeuGlnLeuAenSerLeuGlyAlaAlaAlaAlaAlaGlyThrAlaGly	359
QY	921	-----CCCCGCGCTCTCCCGCGCGCC-----	944
Db	360	AlaAlaGlyThrThrAlaSerLeuLeuLeuSerGluProSerAlaArgProSerPheSer	379
QY	945	CTGCAG-----GCCCTGCTGCGCTCTGCGCGTACGCGCGCGCGCGAGCC	989
Db	380	IleGluAsnIleIleGlyGlyGlyProAlaAlaPro-----GlyGlySerAla	395
QY	990	GGCGCGGCTGGCGCGCGAGCGGAGGCGCACCGACCGCGCGCGCTCTCTGCTGC	1049
Db	396	ValGlyAlaGlyValAlaGlyGlyThrGlyGly-----	406
QY	1050	ACCTCTCCCGCGCGCGCCCCCGCCAGCCACTCCGAGGCGCGCGCGCGCGCGCGC	1106
Db	407	--SerGlyGlyGlySerThrAlaGlnSerPheLeuArgProGlyThrValGlnSer	425
QY	1107	-----GCA	1105
Db	426	AlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArgThrThrAlaThrIleAla	445
QY	1110	CTGTGACTG-----CCCCCTGGCGTCCCGGAGC	1139
Db	446	ProIleLeuSerValProLeuSerGlyGlnPheLeuGlnProAlaAlaSerAlaAlaAla	465
QY	1140	CTGTGCGGCGCTTAGTCCGCGCTCTGCGCGCGCA	1175
Db	466	AlaAlaAlaAlaAlaGlnAlaLysTyrProAla	477
RESULT 9			
ABR63545	ID	ABR63545 standard; protein; 464 AA.	
AC	ABR63545;		
XX	18-SEP-2003	(first entry)	
DT	Danio rerio foxd3	homologue murine HFH2.	
XX	Craniofacial	malformation; variant foxd3; osteopathic; cardiac;	
XX	cytostatic; auditory; neuroprotective; cleft palate syndrome;		
XX	congenital heart disease; cancer; mouse; HPH2.		
OS	Mus musculus.		
XX	WO2003048196-A2.		
XX	12-JUN-2003.		
XX	09-DEC-2002; 2002WO-EP013936.		
PF	07-DEC-2001; 2001EP-00129205.		
XX	(GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.		
XX	Knapik E, Sachdev S;		
XX	WPI; 2003-513737/48.		
XX	New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing		
XX	PT and treating craniofacial malformations, such as cleft palate syndrome,		
XX	loss of middle ear ossicles, congenital heart disease, cancer and mental		
XX	retardation.		
XX	Claim 34: Page 66-67; 70pp; English.		

New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing and treating craniofacial malformations, such as cleft palate syndrome, loss of middle ear ossicles, congenital heart disease, cancer and mental retardation.

Claim 34: Page 66-67: 70pp; English.

XX The present invention provides the protein and coding sequences of Danio
 CC rerio foxd3. These can be used in the diagnosis and treatment of
 CC craniofacial malformations, such as cleft palate syndrome, loss of middle
 CC ear ossicles, congenital heart disease, Nihawa-Kuroki syndrome, cancer
 CC and mental retardation. The present sequence is a homologue of the
 CC zebrafish foxd3 protein, and is murine HFH2

XX SQ Sequence 464 AA;

Alignment Scores:

Pred. No.: 2,426-21 Length: 464
 Score: 466.00 Matches: 154
 Percent Similarity: 39.25% Conservative: 34
 Best Local Similarity: 32.15% Mismatches: 173
 Query Match: 21.29% Indels: 118
 DB: 6 Gaps: 16

US-10-087-080-31 (1-1209) x ABR63545 (1-464)

QY 37 GGGGACACAGGGCGAGTACTGAGGCG-----GCGGGGCGAGCGACGCGCG 87
 DB 5 GlySerGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp 24
 QY 88 TCCCCGCTCTCGCGCGGAGAGACTCCCTG-----GGCTCAGATGGGACTGC 138
 DB 25 IleAspValValGlyGluGlyAspGlyLeuGluGluLysAspSerAspAlaGlyCys 44
 QY 139 -----GCGGCAAGCGC-----TCC 153
 DB 45 SerProAlaGlyProProAspLeuArgLeuAspGluAlaAspGluGlyProValSer 64
 QY 154 GGGGGCGGCGCGCAGAGATACGAGCGCGAGCGGCGGAGGCGGCGGCGGCGG 213
 DB 65 AlaHisGlyGlnSerGlnProGlnAlaLeuAlaLeuProThrGluAlaThrGlyPro 84
 QY 214 GCGCGGAGAGCGGATCCCGCAGCAGCTGTGTCAGCGGTGTGCGGAGGCGCGGAG 273
 DB 85 GlyAsnAspThrGlyAlaProGluAlaAspGlyCysLys---GlyGlyGluAspAlaVal 103
 QY 274 GCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333
 DB 104 ThrGlyGlyGlyProGlyAlaGlySerGlyAlaThrGlyGlyLeuThrProAsnLys 123
 QY 334 CCATATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393
 DB 124 ProLysAsnSerLeuValLysProTyrSerTyrIleAlaLeuIleThrMetAlaIle 143
 QY 394 CCGGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453
 DB 144 LeuGlnSerProGlnLysLysLeuThrLeuSerGlyIleCysGluPheIleSerAsnArg 163
 QY 454 TTCCCTTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513
 DB 164 PheProTyrTyrArgLysPheProAlaThrGlnAsnSerIleArgHisAsnLeuSer 183
 QY 514 CTCACGACTGTTGCTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573
 DB 184 LeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---GlyLysGlyAsn 202
 QY 574 TACTGATGCTCAACCCCAACGCGGAGTACCTTCGCGGCGGCGGCGGCGGCGGCGG 633
 DB 203 TyrTrpThrLeuAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArg 222
 QY 634 CGCAAGCGC----- 642
 DB 223 ArgLysArgPheLysArgHisGlnGlnHisLeuArgGluGlnThrAlaLeuMetMet 242
 QY 643 -----CTCAGCCACCGCGCGGCTCCCGCGCGCGCGCGCGCGCGCGCGG 681
 DB 243 GlnSerPheGlyAlaTyrSerLeuAlaAlaAlaGlyAlaGlyProTyrGlyLeuHis 262
 QY 682 CCGGAGGAGCG 741

DB 263 ProAlaAlaAlaAlaGly-----AlaTyrSerHisProAlaAlaAlaAlaAlaAla 280
 QY 742 CCGCGCATGCGCTCGCGCG-----CCCGCCAG 768
 DB 281 -GlyCysCysGlyAlaProValProValArgAlaThrThrArgGlySerArgAlaAla 300
 QY 769 GAGGAGCGCGCGCG-----CCGCGGCAAGTTCCTCCAGCTCCTTC 810
 DB 300 aArgSerAlaProAlaAlaLeuGlyArgAlaGlyProGlnSerGlyArg-LeuArgLeu 320
 QY 811 GCCATCGACAGCATCTCGCGCAAGCGCTTCGCGCG-----TCGCTCAGG 858
 DB 320 laalaArgSerGluProAlaAlaThrAlaGlnHisProGlyArgArgArgGlySerArg 340
 QY 859 GACACGCGCGCGCGCGCGCTTCAAGTGGCGC-----CGCGCGCTCCCGCGCTG 912
 DB 340 LysHisGlyGlyArgGlyGlyHisValAlaHisGlnValArgAlaGlnCysAlaAlaVal 360
 QY 913 CCGCGTTCCTCGCGCTCTCCCGCGCGCTCGAGGCGCTGCTGCGCTCTGCGCG 972
 DB 360 alValGlnHisArgGluHisArgGlyIleArgAlaProGlyGlySer----- 377
 QY 973 TACGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1032
 DB 378 -----AlaGlyGlyGlyGlySerGlyGlyAlaGlyGlyGlyGlyGly----- 392
 QY 1033 CCGCGCTCTGCTGTCACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092
 DB 393 -----SerGlyGlyGlyGlyAlaGlnSerPheLeuArgProp 406
 QY 1093 GCGCGCGCGCGCG----- 1106
 DB 406 roGlyThrValGlnSerAlaAlaLeuMetAlaThrHisGlnProLeuSerLeuSerArg 426
 QY 1107 -----GCACCTGTACTG-----CCCC 1122
 DB 426 hrThrAlaThrIleAlaProIleLeuSerValProLeuSerGlyGlnPheLeuGlnPro 446
 QY 1123 CTGCGGTCGCGCGCGCGCTCGAGGCGCGCTTAGTCGCGCTGCTCGCGCGCGCA 1175
 DB 446 laalaSerAlaAlaAlaAlaAlaAlaAlaValGlnAlaLysTrpProAla 463

RESULT 10
 ABR63546
 ID ABR63546 standard; protein; 396 AA.
 XX ABR63546;
 AC ABR63546;
 DT 18-SEP-2003 (first entry)
 DE Danio rerio foxd3 homologue chicken CMH3.
 DE Craniofacial malformation; variant foxd3; osteopathic; cardiac;
 KW cytosolic; auditory; neuroprotective; cleft palate syndrome;
 KW congenital heart disease; cancer; chicken; CMH3.
 OS Gallus gallus.
 XX WO2003048196-A2.
 XX 12-JUN-2003.
 XX 09-DEC-2002; 2002WO-EP013936.
 XX 07-DEC-2001; 2001EP-00129205.
 XX (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.
 XX Knapik E, Sachdev S;
 XX WPI; 2003-513737/48.

PT New foxd3 polynucleotide and their (ant)agonists, useful for diagnosing
 PT and treating craniofacial malformations, such as cleft palate syndrome,
 PT loss of middle ear ossicles, congenital heart disease, cancer and mental
 retardation.

XX Claim 34; Page 68-69; 70pp; English.

CC The present invention provides the protein and coding sequences of Danio
 CC rerio foxd3. These can be used in the diagnosis and treatment of
 CC craniofacial malformations, such as cleft palate syndrome, loss of middle
 CC ear ossicles, congenital heart disease, Niikawa-Kuroki syndrome, cancer
 CC and mental retardation. The present sequence is a homologue of the
 CC zebrafish foxd3 protein, and is chicken CNH3

XX Sequence 396 AA;

Alignment Scores:
 Pred. No.: 5,06e-20 Length: 396
 Score: 445.00 Matches: 148
 Percent Similarity: 41.41% Conservative: 40
 Best Local Similarity: 32.60% Mismatches: 115
 Query Match: 20.33% Indels: 22
 DB: 6 Gaps: 21

US-10-087-080-31 (1-1209) x ABR63546 (1-396)

QY 49 GCGAGTACCTGGAGGCG-----GCGGGCGGAGAGCGCGCTCCCGCTGTCG 99
 DB 7 GlySerAspMetSerGlyGlnThrAlaLeuAlaGluAspValAspValVal 26
 QY 100 GCGGCGGAGACGACTCCCTGGGTGATGAGGAGTGGCGGCCAAG---CGTCCGCG 156
 DB 27 GlyGluGlyAspAspAlaProGlyLysAspGlyAspGlyAlaArgSerProAlaAla 46
 QY 157 -----GCGGCGGCGCGC 168
 DB 47 LeuProArgLeuProLeuAspGluAlaGluProGlyLeuProGlyAlaGlyGlu 66
 QY 169 AGATACCGCAGGCGCGGCGGAGCGAGTGGCGGAGCGG-----GCCGGCGCGGAGGA 224
 DB 67 SerGlySerSerGlySerGlySerProAlaProAlaGlyProGluGlyArgGlyGly 86
 QY 225 GCGATCCGCGCAGCAGCTGCTGCGGCTGTGGCGGAGCGCGGAGCGCGCGCGCGC 284
 DB 87 Gly-----GlyGlyGlyGly-GlyGlyGlyGlyGlyGlyAla 98
 QY 285 GGGGCCAGCGCGCGCGCGGCGGAGCGCGAGGCGTGCACGAGCAAGCATATACGCG 344
 DB 98 rGlyGlyAlaAlaAlaAlaAla-----AlaGlyGlnSerLysProLysSer 116
 QY 345 GCGGCCAACGCCCTTACTGTATCGTACATCGCGCTCATCGCATCGCGCATCGCGACTCGCG 404
 DB 116 rLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSer 136
 QY 405 GGGCGGCGCTGACGCTGGCGGAGATCAACAGTACTCTCATGGCAAGTTCCCTTTT 464
 DB 136 oGlnLysLeuThrLeuSerGlyIleCysGluPheIleSerAsnArgPheProTyr 156
 QY 465 CCGCGGCGAGCTACACGGCGCTGGCGCAACTCCGTGCGCCACAACTTCGCTCAACGACTG 524
 DB 156 rArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCy 176
 QY 525 CTTCGTCAAGTGTGGCGGACCCCTCGCGCTCGGGCGGCGAGCAACTACTGATGCT 584
 DB 176 sPheValLysIleProArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLe 195
 QY 585 CAACCCCAACAGCAGTACACTTCGCGCGAGCGGGTCTTCGCGCGCGCGCGCAAGCGCCT 644
 DB 195 uAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPh 215
 QY 645 CAGCCACCGCGCGCGG----- 660
 DB 215 eLysArgHisGlnGlnGluHisLeuArgAspGlnThrAlaLeuMetMetGlnGlyPheG1 235

QY 661 -----GTCCCCCGCGCGCGCTGGCGCGGAGAGCCCGCGGCTCCCGCGCGC 710
 DB 235 yAlaTyrGlyLeuAlaGlyProTyrGlyArgProTyr-----GlyLeuProProG1 252
 QY 711 CCGCGCGCGCGCGCGCGCGCGCGCTCGCGCGCATCGCTCGCGCGCGCGCGCGCG 770
 DB 252 yAlaTyrProHisProAlaAla----- 259
 QY 771 GGAGCGCGCGCGCGCGCGCGCGCGAGTTCCTCCAGTCTCTTCCCATCCACAGCATCTCGC 830
 DB 259 -----ProValGlyProMetLeuProAlaValPro----- 277
 QY 951 GGCCTGCTGCGCTCTCGCGGTACGGCGGCGGAGCGCGCGGCTGGCGCGCGCGCA 1010
 DB 278 -----LeuLeuPro-----SerGlyGluLeuSerArg---LysAlaPheAs 290
 QY 1011 GCGGAGGTGCCACGACGCGCGCGCTCTCTCTTCCACTCTCCCGCGCGCGCGC-- 1068
 DB 290 nAlaGlnLeuGlyProSerLeu---GlnLeuGlnLeuSerSerLeuGlyAlaAlaGlySe 309
 QY 1069 -----CCGCGCAAGCCA-----CTCCGAGG 1088
 DB 309 rIleValLysSerGluProSerSerArgProSerPheSerIleGluAsnIleIleGlyG1 329
 QY 1089 CCGCGCGCGCGCGCGCGCGCTGTACTGCCCC-----CTGCGGCTGCC 1133
 DB 329 yProAlaAlaSerSerAla-----ProSerAlaGlnThrPheLeuArgProPr 345
 QY 1134 GCGACCCCTGACGCGCGCTTACTC---CGNCGTCTCT----- 1167
 DB 345 oValThrValGlnSerGlyLeuValAlaHisGlnProLeuAlaLeuAlaArgThrThrAl 365
 QY 1168 -----GCCCGCGCACCTGTCTGTACCGCGTGGAGACGCTC 1200
 DB 365 aAlaIleAlaProIleLeuSerValProThrAsnIleIle 378
 RESULT 11
 ABUS6559
 ID ABUS6559 standard; protein; 409 AA.
 XX AC ABUS6559;
 XX XX 02-APR-2003 (first entry)
 DE Lung cancer-associated polypeptide #152.
 XX Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antinflammatory; antiathmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX Unidentified.
 OS WO200286443-A2.
 PN 31-OCT-2002.
 XX 18-APR-2002; 2002WO-US012476.
 XX 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.

214	GGCGGCGAGGAGCGCATCCGCGACGAGCTCTGCACGGTGTGTCGAGGCGCGCGAG	273
QY		
Db	:::::	
20	GlyGlyGluGlnSerTyTyArgAlaAalaAala-----AlaGlyGlyTyr	37
QY		
274	GCGGGGCGGGCGGCCGAGCGCG-	297
QY		
38	ThrAlaMetProAlaProMetSerValTySerHisProAlaHisAlaGluInTyPro	57
Db	::::	
298	GCGCGCGCGGGAGCGCGGAGGGTGCCGCGACGATATACGGGGCGGCC---	351
QY		
Db	:::::	
58	GlyGlyMetAlaArgAlaTyGly-----ProTyThrProGlnProGlnPro	73
QY		
352	-----AAGCCCCCTACTCGTACATCGCTCATCGCCATGGCCATTCCGCGAC	399
QY		
74	LysAspMetValLysProProTySerTyfLeAlaLeuileThrMetAlaileGlnaen	93
Db		
400	TGCGGCGCGCGCGCTTACGCTCGGGAGATCAACAGTACTCTCATGGCAGTTCCC	459
QY	::::	
Db	:::::	
94	AlaProAspLysLysIleThrLeuaSngLyfLeTyrgInPheIleMetAspargPhePro	113
QY		
460	TTTTTCCGCGGCATCACCGGGCTGGCGCAACTCCGTGGCGCCACAACTTTTCGTCAAC	519
QY	::::	
114	PheTyArgAspAmLysGlnGlyTrpGlnAsnSerIleargHisAnLeuSerLeuan	133
Db		
520	GACTGCTTCGTCAAAGTGTCTGGCGAACCCCTTCGCGGCCCTGGGGCAAGGACAAC	579
QY		
134	GluCySPheValLysValProArgAspAspLysLysPro---GlyLysGlySerTyTrp	152
Db	::::	
	:::::	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 29, 2004, 11:57:57 ; Search time 20.5 Seconds
(without alignments)
6089.346 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2.69
Sequence: 1 atgaagtggagggttcgt.....tgsagacgtccttagcttga 1209

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 369414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xl
-Q=/cgn2_1/USPTO_spool/p/US10087080/runat_29032004_114527_6585/app_query.fasta_1.1351
-DB=Issued Patents AA -QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pfo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080@cgn2_1 1.27@runat_29032004_114527_6585 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.5	19.7	553	US-09-083-351-2	Sequence 2, Appli
2	431.5	19.7	553	US-09-083-352-2	Sequence 2, Appli
3	394.5	18.0	330	US-09-976-594-928	Sequence 928, App
4	387	17.7	473	US-08-857-076-99	Sequence 99, Appl
5	347	15.9	480	US-07-882-292-2	Sequence 2, Appli
6	347	15.9	480	US-08-331-644-2	Sequence 2, Appli
7	347	15.9	480	PCT-US93-04102-2	Sequence 100, App
8	336	15.3	347	US-08-857-076-100	Sequence 14, Appl
9	333	15.2	106	US-09-083-351-14	Sequence 14, Appl
10	333	15.2	106	US-09-083-352-14	Sequence 14, Appl
11	332	15.2	106	US-09-083-351-15	Sequence 15, Appl
12	332	15.2	106	US-09-083-352-15	Sequence 15, Appl

13	329	15.0	544	4	US-09-087-134-14	Sequence 14, Appl
14	320.5	14.6	106	3	US-09-083-351-12	Sequence 12, Appl
15	320.5	14.6	106	3	US-09-083-352-12	Sequence 12, Appl
16	320.5	14.6	2294	4	US-09-252-991A-17231	Sequence 17231, A
17	319	14.6	365	3	US-09-113-309-2	Sequence 2, Appli
18	319	14.6	365	3	US-09-521-109-2	Sequence 2, Appli
19	319	14.6	365	4	US-09-562-332-2	Sequence 2, Appli
20	317	14.5	369	4	US-09-252-991A-25394	Sequence 25394, A
21	315	14.4	663	4	US-09-252-991A-30843	Sequence 30843, A
22	314.5	14.4	106	3	US-09-083-351-13	Sequence 13, Appl
23	314.5	14.4	106	3	US-09-083-352-13	Sequence 13, Appl
24	312.5	14.3	783	4	US-09-252-991A-18035	Sequence 18035, A
25	311.5	14.2	866	4	US-09-252-991A-20509	Sequence 20509, A
26	311	14.2	863	4	US-09-252-991A-26099	Sequence 26099, A
27	307	14.0	838	4	US-09-252-991A-27068	Sequence 27068, A
28	307	14.0	957	4	US-09-252-991A-20408	Sequence 20408, A
29	307	14.0	1706	4	US-09-252-991A-31760	Sequence 31760, A
30	305.5	14.0	106	3	US-09-083-351-7	Sequence 7, Appli
31	305.5	14.0	106	3	US-09-083-352-7	Sequence 7, Appli
32	305.5	14.0	534	4	US-09-087-134-11	Sequence 11, Appl
33	304.5	13.6	446	4	US-09-252-991A-27110	Sequence 27110, A
34	304.5	13.9	518	3	US-09-113-309-19	Sequence 19, Appl
35	304.5	13.9	518	4	US-09-562-332-19	Sequence 19, Appl
36	304.5	13.9	518	4	US-09-562-332-19	Sequence 19, Appl
37	303	13.8	467	4	US-09-252-991A-18296	Sequence 18296, A
38	303	13.8	586	4	US-09-252-991A-24514	Sequence 24514, A
39	303	13.8	720	4	US-09-252-991A-21881	Sequence 21881, A
40	302.5	13.8	726	4	US-09-252-991A-20675	Sequence 20675, A
41	302	13.8	904	4	US-09-976-594-615	Sequence 615, App
42	301.5	13.4	595	4	US-09-370-838-187	Sequence 187, App
43	298	13.6	809	4	US-09-252-991A-31759	Sequence 31759, A
44	297.5	13.2	467	4	US-09-252-991A-18296	Sequence 18296, A
45	296.5	13.5	495	4	US-09-252-991A-31949	Sequence 31949, A

ALIGNMENTS

RESULT 1

US-09-083-351-2
; Sequence 2, Application US/09083351

; Patent No. 6087107

; GENERAL INFORMATION:

; APPLICANT: Sheffield, Val C.

; APPLICANT: Alward, Wallace L.M.

; APPLICANT: Stone, Edwin M.

; APPLICANT: Nishimura, Darryl

; APPLICANT: Patil, Shiva

; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR

; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN

; TITLE OF INVENTION: TRANSCRIPTION FACTOR

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/083,351

; FILING DATE: 22-MAY-1998

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Attold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: UIA-029.02

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

```

272 rProProGlySerLeuProSerAlaArgProLeuSerLeuAspGlyAlaAspSerAlaPr 299
895 GCACCCTGCCCGCGCGTCCCGGTTCCCGCGCTCCTCCC----- 935
292 oProPro-ProAla-----ProSerAlaProProProHisHisSerGlnGlyP 308
936 -----CGCGCGCGCCTGCAGGGCCCTGCTCCCGCTC 966
308 heSerValAspAsnIleMetThr-SerLeuArgGlySerProGlnSerAlaAlaGluL 328
967 TCGCGGTACGGCGCGGAGCGCGCTGGCGCGCGCGAGGCGGAGGTGCCACCG 1026
328 eu-----SerSerGlyLeuLeuAlaSerAlaAlaSerS 340
1027 ACCGCGCC-----GCCCTCCCT----- 1043
340 exArgAlaGlyIleAlaProProLeuAlaLeuGlyAlaTyrSerProGlyGlnSerSerL 360
1044 -----CCTGCACCTCTCCCGCGCGCGCGCGCCCGCC 1074
360 euTyrSerSerProCysSerSerGlnThrSerSerAlaGlySerSerGlyGlyGlyGlyG 380
1075 AAGCCACTCCGAGCGCGCGCGCGCGCGCACCTGTAC---TGCCTCCCTCGG--- 1128
380 lyAlaGlyAlaAlaGlyGly-AlaGlyGlyAlaGlyThrTyrHisCysAsnLeuGlnAla 399
1129 -----CTGCCCGCAGCCCTG 1143
400 MeSerLeuTyrAlaAlaGlyGlyUArgGlyGlyHisLeuGlnGlyAlaProGlyGlyAla 419
1144 CAGCGCGCCTTAGTCCGNCCTCTGCGCGCGCACCTGTCTGACCGGTGGAGACG 1197
420 GlyGlySerAlaValAspAsnProLeuProAspTyrSerLeuProProValThr 437

RESULT 2
US-09-083-352-2
; Sequence 2, Application US/09083352
; Patent No. 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELLIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Artold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids

```

Qy	286	GGGCCAGCGCGGGCGGGCGGGAGCGAGCGAGCGATATACGCGG	345
Db	8	GlyProGlyAla-	15
		-----ProThrThrGlu	
Qy	346	CGGCCCAAGCCCCCTACTCGTACATCGCGCTCATGCCATGCCATCCGCACTCGCGG	405
Db	16	ProThrLysProProTySerTyIleAlaLeuIleAlaMetAlaIleGlnSerPro	35

Qy	406	GGCGGCGCGTTGACGCTCGCGGAGATCAACAGGTACCTCATGGCGAAGTCCCTCTTTTC	465

```

Db      36 GlyGlnArgAlaThrLeuSerGlyIleTyrArgTyrIleMetGlyArgPheAlaPheTyr 55
QY      466 CGCGGCGAGCTACACGCGGCTGGCGCACTCCGCGCCACAACTTTCCTCAACGACTGC 525
Db      56 ArgHisAsnArgProGlyTyrPheGlnSerIleArgHisAsnLeuSerLeuAsnGluCys 75
QY      526 TTCGTCAAGTCTGCGGACCCCTCGCGGCGCTGGGCAAGAGCAACTACTGATGACTC 585
Db      76 PheValIleValProArgAspArgTyrPro---GlyLysGlySerTyrTrpThrLeu 94
QY      586 AACCCCAACAGCAGGATACACTTCCCGCAGCGGGTCTTCGCGCGCGCGCGCAAGCGCTC 645
Db      95 AspProAspCysHisAspMetPheGluHisGlySerPheLeuArgArgArgArgPhe 114
QY      646 AGCCAC-----CGCGCGCGTCCCGCG-----CCCGGG 675
Db      115 ThrArgGlnThrGlyAlaGluGlyThrArgGlyProAlaLysAlaArgArgGlyProLeu 134
QY      676 CTGCGGCGCGAGAGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db      135 ArgAlaThrSerGlnAspProGlyValProAsnAlaThrThrGlyArgGlnCysSerPhe 154
QY      712 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
Db      155 ProProGluLeuProAspProLysGlyLeuSerPheGlyLeuValGlyAlaMetPro 174
QY      736 GCCTCG-----CCCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774
Db      175 AlaSerMetCysProAlaThrThrAspGlyArgProArgProMetGluProLysGlu 194
QY      775 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834
Db      195 IleSerThrPro-----Lys 199
QY      835 CCTTTCGCGAGCGCTCGCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
Db      200 Pro-----AlaCysProGlyLeuProValAlaThrSer 211
QY      892 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
Db      212 SerSerSerCysProAlaPhe---GlyPheProAlaGlyPheSerGluAlaGluSerPhe 230
QY      934 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
Db      231 AsnLysAlaProThrProValLeuSerProGluSerGlyIleGlySerSerTyrGlnCys 250
QY      994 CGGCTG----- 999
Db      251 ArgLeuGlnAlaLeuAsnPheCysMetGlyAlaAspProGlyLeuGluHisLeuLeuAla 270
QY      1000 GCGCGCGCGAGCGAGGTCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
Db      271 SerAlaAlaProSerProAlaProProThrProProGlySerLeuArgAlaProLeuPro 290
QY      1060 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
Db      291 -----LeuProThrAspHisGlyGluProTrpValAlaGlyPheProValGlnGly 308
QY      1120 CCCCTGCGCGTCCCGCGAGCCCTGCAGCGCGCGCTTAGTCCGCGCGTCTCTGCG 1170
Db      309 GlySerGlyTyrProLeuGlyLeuThrProCysLeuTyrArgThrProGly 325

```

RESULT 4

US-08-857-076-99

; Sequence 99, Application US/08857076C

; Patent No. 6225120

; GENERAL INFORMATION:

; APPLICANT: Ruvkun, Gary

; APPLICANT: Kimura, Koutarou

; APPLICANT: Patterson, Garth

; APPLICANT: Ogg, Scott

; APPLICANT: Paradis, Suzanne

```

; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-857-076-99

```

Alignment Scores:

```

Pred. No.: 5,77e-16 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 3 Gaps: 15

```

US-10-087-080-31 (1-1209) x US-08-857-076-99 (1-473)

```

QY      70 GCGCGCAGCGAGCGCGCTCCCGCTCGCGCGCGCGGAGCAGCTCCCTGGGCTCAGAT 129
Db      88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
QY      130 GGGGACTGCGCGGCGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189
Db      107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnAlaAla 222
QY      190 GAACAGAGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
Db      123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
QY      250 GCGGTGTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
Db      143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
QY      310 AGCGCGAGGAGGTGCACGACGACCAAGCCATATACCGCGCGCGCGCGCGCGCGCGCG 369
Db      160 AlaLysThrPheLysArgSerTyrProHisAla-----LysProTyrSerTyr 176
QY      370 ATCCGCTCATGCCATGCGCATCCGCGACTCCGCGCGCGCGCGCGCGCGCGCGCGCG 429
Db      177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
QY      430 ATCAACGAGTACTCATGGCAAGTTCCCTTTTCGCGCGCGCGCGCGCGCGCGCGCGCG 489
Db      197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTrpGln 216
QY      490 AACTCCGTGCGGCACACCTTTCGCTCAACGACTGCTGCTCAAGTGTCTGCGCGCGCGCG 549
Db      217 AsnSerIleArgHisLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
QY      550 TCGCGGCGCTGGGCGAAGCAGCACTACTGATGCTCAACCCCAACAGCGAGTACACCTTC 609
Db      237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
QY      610 GCGACGCGGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db      256 GluAsnGlyCysTyrLeuArgArgGlnLysArg----- 266
QY      670 CCCGCGCTGCGCGCGCGAGGAGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCG 729
Db      267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
QY      730 GCGCGCGCTCGCGCGCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
Db      278 GlyGlyGlySerGlySerGlySerGlyAlaLysGlyGlyGlyGlyGlySerArgLysAsp 297

```

```

QY 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGCAGCATCTGGCAGCCCTTCCGC 843
Db 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGCAGCATCTGGCAGCCCTTCCGC 843
QY 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGCAGCATCTGGCAGCCCTTCCGC 843
Db 784 CCGCGGGCAAGTCTCCAGCTCTTCCGATCGCAGCATCTGGCAGCCCTTCCGC 843
QY 844 AGCGTCCCTCAGGACAGCGCCCGGACAGCCTTTCAGTGGGGCGCGCCCTGC 903
Db 844 AGCGTCCCTCAGGACAGCGCCCGGACAGCCTTTCAGTGGGGCGCGCCCTGC 903
QY 844 AGCGTCCCTCAGGACAGCGCCCGGACAGCCTTTCAGTGGGGCGCGCCCTGC 903
Db 844 AGCGTCCCTCAGGACAGCGCCCGGACAGCCTTTCAGTGGGGCGCGCCCTGC 903
QY 904 CCGCGCTGCGCGGTTCCCGGCTCTCCCGCGCGCCCTTCCGAGGCGCTGCTCG 963
Db 904 CCGCGCTGCGCGGTTCCCGGCTCTCCCGCGCGCCCTTCCGAGGCGCTGCTCG 963
QY 904 CCGCGCTGCGCGGTTCCCGGCTCTCCCGCGCGCCCTTCCGAGGCGCTGCTCG 963
Db 904 CCGCGCTGCGCGGTTCCCGGCTCTCCCGCGCGCCCTTCCGAGGCGCTGCTCG 963
QY 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
Db 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
QY 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
Db 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
QY 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
Db 964 CTCTGCGGTACGCGGCGGCGAGCGCGGCTGGCGCGCGCGCGAGCGCGAGTGCA 1023
QY 1024 CCGACCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1062
Db 1024 CCGACCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1062
QY 1063 GCGCGCCCGCCCAAGCCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 1063 GCGCGCCCGCCCAAGCCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
QY 1063 GCGCGCCCGCCCAAGCCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 1063 GCGCGCCCGCCCAAGCCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
QY 1123 CTGCGGCTGCGCGCGCGCTGCGAGCGCGCTTAGTCGCGCGCTGCGCGCGCG 1182
Db 1123 CTGCGGCTGCGCGCGCGCTGCGAGCGCGCTTAGTCGCGCGCTGCGCGCGCG 1182
QY 1183 TAC 1185
Db 1183 TAC 1185
QY 397 Phe 397
Db 397 Phe 397

RESULT 5
US-07-882-292-2
; Sequence 2, Application US/07882292
; Patent No. 5324638
; GENERAL INFORMATION:
; APPLICANT: Tao, Mufan
; APPLICANT: Lai, Eseng
; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: c/o Cooper and Dunham, 30 Rockefeller
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/882,292
; FILING DATE: 19920513
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41472
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-977-9550
; TELEFAX: 212-664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

```

MOLECULE TYPE: protein

US-07-882-292-2

Alignment Scores:

```

Pred. No.: 1,41e-13 Length: 480
Score: 347.00 Matches: 124
Percent Similarity: 40.24% Conservative: 41
Best Local Similarity: 30.24% Mismatches: 144
Query Match: 15.85% Indels: 101
DB: 1 Gaps: 17

```

US-10-087-080-31 (1-1209) x US-07-882-292-2 (1-480)

```

QY 22 CTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
Db 22 CTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
QY 81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly- 95
Db 81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly- 95
QY 82 GCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 117
Db 82 GCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 117
QY 96 ---ProGlnProLeuLeuLeuProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114
Db 96 ---ProGlnProLeuLeuLeuProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114
QY 118 CTGGCTCAGATGGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177
Db 118 CTGGCTCAGATGGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 177
QY 115 LeuGlyAlaLysGlyGlu- 130
Db 115 LeuGlyAlaLysGlyGlu- 130
QY 178 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237
Db 178 CAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237
QY 131 ValGlyProAspGluLysGlyAlaGlyAlaGlyGlyGlu- 146
Db 131 ValGlyProAspGluLysGlyAlaGlyAlaGlyGlyGlu- 146
QY 238 GCAGCTGTGACGCGGTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db 238 GCAGCTGTGACGCGGTGTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
QY 147 ---LysLysGlyAlaGlyGlyLys 155
Db 147 ---LysLysGlyAlaGlyGlyLys 155
QY 298 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
Db 298 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
QY 156 AspGlyGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 173
Db 156 AspGlyGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 173
QY 358 CCTACTCTGATCTGCGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
Db 358 CCTACTCTGATCTGCGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
QY 174 ProPheThrThrAsnAlaLeuLeuMetMetAlaAlaArgGlnSerProGluLysArgLeu 193
Db 174 ProPheThrThrAsnAlaLeuLeuMetMetAlaAlaArgGlnSerProGluLysArgLeu 193
QY 418 AGCTGGCGGAGATCAAGAGTACCTCATGGCAAGTTCCCTTTTCCGCGCGCGAGTAC 477
Db 418 AGCTGGCGGAGATCAAGAGTACCTCATGGCAAGTTCCCTTTTCCGCGCGCGAGTAC 477
QY 194 ThrLeuAsnGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 213
Db 194 ThrLeuAsnGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 213
QY 478 AGCGCTGGCGCAACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
Db 478 AGCGCTGGCGCAACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
QY 214 GlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233
Db 214 GlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233
QY 538 CTGCGCGACCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Db 538 CTGCGCGACCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
QY 234 ProArgHisTyrAspAspPro---GlyLysGlySerThrMetLeuAspProSerSer 252
Db 234 ProArgHisTyrAspAspPro---GlyLysGlySerThrMetLeuAspProSerSer 252
QY 598 GAGTACACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
Db 598 GAGTACACTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
QY 253 AspAspValPheIleGlyGlyThrThrGlyLysLeuArgArgArgSerThrThrSerArg 272
Db 253 AspAspValPheIleGlyGlyThrThrGlyLysLeuArgArgArgSerThrThrSerArg 272
QY 643 ---CTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
Db 643 ---CTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
QY 273 AlaLysLeuAlaPheLysArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292
Db 273 AlaLysLeuAlaPheLysArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292
QY 691 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
Db 691 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 723
QY 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312
Db 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312
QY 724 ---CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 724 ---CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
QY 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332
Db 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332
QY 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
QY 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnAsnHisSerPheSerThrAlaAsnGly 352
Db 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnAsnHisSerPheSerThrAlaAsnGly 352

```


QY 907 -----CGCGTGC-----CGCGTTC 921
 Db 393 LeuAsnPro-CysSerValAsnLeuLeuAlaGlyGlnThrSerTyrPheProHisVa 412
 QY 922 CCGCGCTCCTCCCGCGCGCCCTGCAGGCGCCCTGCTGCGCTCTGCGGTACGCGCG 981
 Db 412 1ProHisProSer-----MetThrSerG1 420
 QY 982 GCGAGCGCGCGCGTGGCGCGCGGAGCGGAGCGGAGTGCACCGACCGCGCGCCCTC 1041
 Db 420 nThrSerThrSerMetSerAlaArgAlaAlaSerSerThrSerProGlnAlaProSe 440
 QY 1042 CTGCTTTGCACCTCTCCCGCGCGCGCC 1069
 Db 440 rThrLeuProCysGluSerLeuArgPro 449

RESULT 7

PCT-US93-04102-2
 ; Sequence 2, Application PC/TUS9304102
 ; GENERAL INFORMATION:
 ; APPLICANT: Tao, Wufan
 ; TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White
 ; STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/04102
 ; FILING DATE: 19930430

CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/882,292
 ; FILING DATE: 13-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 41472A-PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-977-9550
 ; TELEFAX: 212-664-0525
 ; TELEX: 422523 COOP UI
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 480 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

PCT-US93-04102-2

Alignment Scores:
 Pred. No.: 1,41e-13 Length: 480
 Score: 347.00 Matches: 124
 Percent Similarity: 40.24% Conservative: 41
 Best Local Similarity: 30.24% Mismatches: 144
 Query Match: 15.85% Indels: 101
 DB: 5 Gaps: 17

US-10-087-080-31 (1-1209) x PCT-US93-04102-2 (1-480)

QY 22 CTCGCGCGCGCCACCGGACAGAGCGGCGTGCCTGGAGGCGCGGCGCGACGAC 81
 Db 81 ProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly----- 95

QY 82 GCGCGCTCCCGCTG-----TCGGCGCGCGGAGAC-----GACTCC 117
 Db 96 ---ProGlnProLeuLeuLeuProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114
 QY 118 CTGGCTCAGATGGGAGACTGCGCGGCAAGCCCTCGCGGCGCGGCGCGGCGGAGATACG 177
 Db 115 LeuGlyAlaLysGlyGlu-----ProGlyGlyGlyProAlaGluLeuAlaPro 130
 QY 178 CAGGCGACGCGCAACAGAGTGGCGGAGGCGCGCGCGCGGAGGAGGAGGATCCCGGCA 237
 Db 131 ValGlyProAspGluLysGlyAlaGlyAlaGlyGlyGlu----- 146
 QY 238 GCAGCTCTGCAGCGGTGTGGCGGAGGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGG 297
 Db 147 -----LysLysGlyAlaGlyGlyGlyLys 155
 QY 298 GCGCGCGCGGAGCGCGGAGGTGCACGAGCAAGCCATATACGCGGCGCGCCCAAGCCC 357
 Db 156 AspGlyGluGlyLysGlyGly-----AspLysAsnAsnGlyLysTyrGluLysPro 173
 QY 358 CCTTACTCGTACATCGCGCTCATCGCATCGCATCGCGACTCGCGGCGCGGCGGCTTG 417
 Db 174 ProPheThrTyrAsnAlaLeuMetMetAlaLeuMetMetAlaLeuSerProGluLysArgLeu 193
 QY 418 AGCTGCGGAGATCAACAGTACTCTCATCGGCAAGTCCCTTTTCGCGGCGGAGTAC 477
 Db 194 ThrLeuAsnGlyTyrGluPheMetMetLysAsnPheProTyrTyrArgGluAsnLys 213
 QY 478 ACGGCTGGCGCAACTCCGTCGCGCCACAACTTCGCTCAACGACTCTTCGTCACAGGTG 537
 Db 214 GlnGlyTyrGlnAsnSerIleArgHisAsnLeuSerLeuAsnLysCysPheValLysVal 233
 QY 538 CTGCGGACCCCTCGCGGCGCTGGGCAAGGACAACTACTGGATGCTCAACCCCAAGCAGC 597
 Db 234 ProArgHisTyrAspAspPro---GlyLysGlyAsnTyrTrpMetLeuAspProSerSer 252
 QY 598 GAGTACACCTTCGCGGCGCGGTCTTCGCGCGCGCGCGCAAGCGC----- 642
 Db 253 AspAspValPheIleGlyGlyThrThrGlyLysLeuArgArgSerThrThrSerArg 272
 QY 643 -----CTCAGCCACCGCGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGGAG 690
 Db 273 AlaLysLeuAlaPheLysArgArgAlaArgLeuThrSerThrGlyLeuThrPheMetAsp 292
 QY 691 GCCCGCGGC-----CTCCCGCGCGCGCGC-----CGCGCGCGG 723
 Db 293 ArgAlaGlySerLeuTyrTrpProMetSerProPheLeuSerLeuHisHisProArgAla 312
 QY 724 -----CCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
 Db 313 SerSerThrLeuSerTyrAsnGlyThrThrSerAlaTyrProSerHisProMetProTyr 332
 QY 754 TCGCGCGCGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
 Db 333 SerSerValLeuThrGlnAsnSerLeuGlyAsnHisSerPheSerThrAlaAsnGly 352
 QY 808 TTCGCCATCGACAGCATCTCGCGCAAG-----CCCTTCGCGAGCGCGTCCGCTCAGGAGC 861
 Db 353 LeuSerValAspArgLeuValAsnGlyGluIleProTyrAlaThrHisHisLeuThrAla 372
 QY 862 ACGCGCGCGCGGACGAGCTTCAGTGGGC-----CGCGCGCGCGCGCGCGCG 906
 Db 373 AlaAlaLeuAlaAlaSerValProCysGlyLeuSerValProCysSerGlyThrTyrSer 392
 QY 907 -----CGCGTGC-----CGCGTTC 921
 Db 393 LeuAsnPro-CysSerValAsnLeuLeuAlaGlyGlnThrSerTyrPhePheProHisVa 412
 QY 922 CCGCGCTCTCCCGCGGCGCGCTGCTGCGGCGCGCTGCTGCGGCGCGGCGCGGCGG 981
 Db 412 1ProHisProSer-----MetThrSerG1 420

Best Local Similarity: 63.00% Mismatches: 19
 Query Match: 15.21% Indels: 2
 DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-351-14 (1-106)

QY 343 CGCGGCCCC---AAGCCCCCTACTCGTACATCGGCTCATCGCCATCGCCATCGCGGAC 399
 DB 2 ArgAArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
 QY 400 TCGCGGCGGCGGCTGACGCTGCGGAGATCAACGAGTACCTCATCGGCAAGTTCGCC 459
 DB 22 SerPProThrLysArgLeuThrLeuSerGluLeuTyrGlnPheLeuGlnSerArgPhePro 41
 QY 460 TTTTTCGCGGCGAGTACACAGGCTGCGGCACTCCGTCGCCCAACCTTTCGCTCAAC 519
 DB 42 PhePheArgGlySerTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61
 QY 520 GACTGCTTCGTCAGGTGCTGCGGACCCCTCGCGGCTGCGGCAAGCAACTACTGG 579
 DB 62 GlucysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
 QY 580 ATGCTCAACCCCAACAGCGAGTACACCTTCGCGGCGGCTTCGCGCGCGCGCGCAAG 639
 DB 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArgArg 100

RESULT 10

US-09-083-352-14

; Sequence 14, Application US/09083352
 ; Patent No. 6207450

; GENERAL INFORMATION:

; APPLICANT: Sheffield, Val C.

; APPLICANT: Alward, Wallace L.M.

; APPLICANT: Stone, Edwin M.

; APPLICANT: Nishimura, Darryl

; APPLICANT: Patil, Shiva

; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS

; TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELLIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/083,352

; FILING DATE: 22-MAY-1998

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Arnold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: UIA-029.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 106 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-083-352-14

Alignment Scores:

Pred. No.:

9.13e-13

Length:

106

Score: 333.00 Matches: 63
 Percent Similarity: 79.00% Conservative: 16
 Best Local Similarity: 63.00% Mismatches: 19
 Query Match: 15.21% Indels: 2
 DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-352-14 (1-106)

QY 343 CGCGGCCCC---AAGCCCCCTACTCGTACATCGGCTCATCGCCATCGCCATCGCGAC 399
 DB 2 ArgAArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21
 QY 400 TCGCGGCGGCGGCTGACGCTGCGGAGATCAACGAGTACCTCATCGGCAAGTTCGCC 459
 DB 22 SerPProThrLysArgLeuThrLeuSerGluLeuTyrGlnPheLeuGlnSerArgPhePro 41
 QY 460 TTTTTCGCGGCGAGTACACAGGCTGCGGCACTCCGTCGCCCAACCTTTCGCTCAAC 519
 DB 42 PhePheArgGlySerTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61
 QY 520 GACTGCTTCGTCAGGTGCTGCGGACCCCTCGCGGCTGCGGCAAGCAACTACTGG 579
 DB 62 GlucysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTrp 80
 QY 580 ATGCTCAACCCCAACAGCGAGTACACCTTCGCGGCGGCTTCGCGCGCGCGCGCAAG 639
 DB 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArgArg 100

RESULT 11

US-09-083-351-15

; Sequence 15, Application US/09083351

; Patent No. 6087107

; GENERAL INFORMATION:

; APPLICANT: Sheffield, Val C.

; APPLICANT: Alward, Wallace L.M.

; APPLICANT: Stone, Edwin M.

; APPLICANT: Nishimura, Darryl

; APPLICANT: Patil, Shiva

; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR

; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN

; TITLE OF INVENTION: TRANSCRIPTION FACTOR

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELLIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/083,351

; FILING DATE: 22-MAY-1998

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Arnold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: UIA-029.02

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 106 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-083-351-15

MOLECULE TYPE: protein

US-09-083-352-15

Alignment Scores:

Pred. No.: 1.05e-12 Length: 106

Score: 332.00 Matches: 62

Percent Similarity: 79.00% Conservative: 17

Best Local Similarity: 62.00% Mismatches: 19

Query Match: 15.17% Indels: 2

DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-351-15 (1-106)

Qy 343 CGCGCGCC---AAGCCCCCTACTCGTACATCGCGCTCATCGCCATCGCGAC 399

Db 2 ArgArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21

Qy 400 TCGGCGGGCGGCGCTTACGCTGGCGGAGATCAACAGTACTCTCATGGCAAGTTCGCC 459

Db 22 SerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhePro 41

Qy 460 TTTTTCGGCGGAGCTACACGGGCTGGCGCAACTCGTGGCGGCAACACTTTCGCTCAAC 519

Db 42 PhePheArgGlyAlaTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61

Qy 520 GACTGCTTGTCAAGTGTGCGCGACCCCTCGCGCCCTCGCGGCAAGCAACTACTCGG 579

Db 62 GluCysPheIleLysLeuProLysGlyLeuGlyArgPro--GlyLysGlyHisTyrTrp 80

Qy 580 ATGCTCAACCCCAACAGCAGTACACTTCGCGGAGGGGTCTTCGCGCGCGCGCAAG 639

Db 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArg 100

RESULT 12

US-09-083-352-15

Sequence 15, Application US/09083352

Patent No. 6207450

GENERAL INFORMATION:

APPLICANT: Sheffield, Val C.

APPLICANT: Alward, Wallace L.M.

APPLICANT: Stone, Edwin M.

APPLICANT: Nishimura, Darryl

APPLICANT: Patil, Shiva

TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS

TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG & ELIOT LLP

STREET: One Post Office Square

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,352

FILING DATE: 22-MAY-1998

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Arnold, Beth E.

REGISTRATION NUMBER: 35,430

REFERENCE/DOCKET NUMBER: UIA-029.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-832-1000

TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 106 amino acids

TYPE: amino acid

STRANDEDNESS: linear

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-083-352-15

Alignment Scores:

Pred. No.: 1.05e-12 Length: 106

Score: 332.00 Matches: 62

Percent Similarity: 79.00% Conservative: 17

Best Local Similarity: 62.00% Mismatches: 19

Query Match: 15.17% Indels: 2

DB: 3 Gaps: 2

US-10-087-080-31 (1-1209) x US-09-083-351-15 (1-106)

Qy 343 CGCGCGGCC---AAGCCCCCTACTCGTATACATCGCGCTCATCGCCATCGCCGAC 399

Db 2 ArgArgProGluLysProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSer 21

Qy 400 TCGGCGGGCGGCGCTTACGCTGGCGGAGATCAACAGTACTCTCATGGCAAGTTCGCC 459

Db 22 SerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhePro 41

Qy 460 TTTTTCGGCGGAGCTACACGGGCTGGCGCAACTCGTGGCGGCAACACTTTCGCTCAAC 519

Db 42 PhePheArgGlyAlaTyrGlnGlyTrpLysAsnSerValArgHisAsnLeuSerLeuAsn 61

Qy 520 GACTGCTTGTCAAGTGTCTGGCGGACCCCTCGCGGCCCTGGCGGCAAGCAACTACTGG 579

Db 62 GluCysPheIleLeuLysLeuProLysGlyLeuGlyArgPro--GlyLysGlyHisTyrTrp 80

Qy 580 ATGCTCAACCCCAACAGCAGATACACTTTCGGCGGAGGGGTCTTCGGCGGCGCGCAAG 639

Db 81 ThrIleAspProAlaSerGluPheMetPheGluAsnGlySerPheArgArgArg 100

RESULT 12

US-09-083-352-15

Sequence 15, Application US/09083352

Patent No. 6207450

GENERAL INFORMATION:

APPLICANT: Sheffield, Val C.

APPLICANT: Alward, Wallace L.M.

APPLICANT: Stone, Edwin M.

APPLICANT: Nishimura, Darryl

APPLICANT: Patil, Shiva

TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS

TITLE OF INVENTION: BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG & ELIOT LLP

STREET: One Post Office Square

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,352

FILING DATE: 22-MAY-1998

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Arnold, Beth E.

REGISTRATION NUMBER: 35,430

REFERENCE/DOCKET NUMBER: UIA-029.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-832-1000

TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 106 amino acids

TYPE: amino acid

STRANDEDNESS: linear

TOPOLOGY: linear

```

281 uProIleTyThrProAsnValValMetProLeuAlaProProThrSerCysProGI 301
966 CTGCGGCTACGGCGCGCGGAGCCGCGCGGCTGGCGCGCGAGCCGAGGTGCCACC 1025
301 nCysPro-----SerThrProAlaTyTrpGlyVal-----AlaProGI 315
1026 GACC---GGCGCGCCCTCTCTGTGTGCACCTCTCCCGGCGGCGGCCCGCCCAAGCCA-- 1080
315 uThrArgGlyProProGlyLeuLeuCysAspLeuAsnAlaLeuPheGlnGlyValProPr 335
1081 -----CTCCGAGGCCCGCGCGGC----- 1119
335 oAsnLysSerIleTyzAspValTrpValSerHisProArgAspLeuAlaAlaProGlyPr 355
1098 CGCGCGCGCGCACCTGTACTGC----- 1119
355 oGlyTrpLeuLeuSerTrpCysSerLeu**GlySer**AspArgGlyArgSerSerLe 375
1120 -CCCTCTCGGCTCCCGCGAGCCCTGCGAGCGCGCTTAGTCCGNCGT 1164
375 uProLeuProProProCys**GlnGlyAlaLysAlaArgArg 390

RESULT 14
US-09-083-351-12
; Sequence 12, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-351-12

Alignment Scores:
Pred. No.: 5 09e-12 Length: 106
Score: 320.50 Matches: 61
Percent Similarity: 71.43% Conservative: 14
Best Local Similarity: 58.10% Mismatches: 29

```

```

Query Match: 14.64%      Indels: 1
DB:          3           Gaps: 1

```

US-10-087-080-31 (1-1209) x US-09-083-351-12 (1-106)

340	QY	ACGGCGGGCCCAAGCCCCCTACTCTGTACATCCGCTCATCGCCATGCGCATCCGCGAC	399
2	Db	ThrArgLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGln	21
400	QY	TCGGGGGGGGCGGCTTCACGCTGGCGGAGATCAACGAGTACTCTCATGGGCAAGTTCGCC	459
22	Db	SerProLysLysArgLeuThrLeuSerGluIleCysGluPheIleSerGlyArgPhePro	41
460	QY	TTTTTTCGGGGCAGCTACACGGGTGGCGGAACCTCCGTGGCCACCAACTTTCGCTCAAC	519
42	Db	TyrTyrArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAen	61
520	QY	GACTCGCTTCGTCAAGAGTGCTCGGACCCCTCGGGGCCCTGGGGCAAGGACAACACTCG	579
62	Db	AspCysPheValLysIleProArgGluProGlyAsnPro--GlyLysGlyAsnTyrTrp	80
580	QY	ANGCTCAACCCCAACAGGAGTACACCTTCGCCACGCGGGTCTTCGCGCGCGCCGCAAG	639
81	Db	ThrLeuAspProGluLysAlaAspMetPheAspAsnGlySerPheLeuArgArgGlyLys	100
640	QY	CGCTCAAGCCACCCG	654
101	Db	ArgPheLysArgGln	105

RESULT 15

```

US-09-083-352-12
; Sequence 12, Application US/09083352
; Patent No 6207450
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
; BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSER: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083.352
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-083-352-12

```

Alignment Scores:

Pred. No.:	5,098-12	Length:	106
Alignment Scores:		Matches:	61
Score:	320.50	Conservative:	14
Percent Similarity:	71.43%	Mismatches:	29
Best Local Similarity:	58.11%	Indels:	1
Query Match:	14.64%	Gaps:	1
DB:	3		

US-10-087-080-31 (1-1209) x US-09-083-352-12 (1-106)

340	QY	ACGGCGGGCCCAAGCCCCCTACTCTAGATCGCGCTCATGCCCATGGCCATCCCGAC	399
2	Db	ThrArgLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGln	21
400	QY	TCGGCGGGCGGGCTTGACGCTGGCGGAGATCAACGAGTACTCATGGCGAAGTTTCCCC	459
22	Db	SerProLysLysArgLeuThrLeuSerGluIleCysGluPheIleSerGlyArgPhePro	41
460	QY	TTTTTCCGGGCGGACTACACGGGTGGCGCACTCGGTGGCGGCACACCTTTCGCTCAAC	519
42	Db	TyrTyrArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn	61
520	QY	GACTGCTTCGTCAAGGTGCTCGGACCCCTCGCGGCCCTGGGGCAAGCAACTACTGG	579
62	Db	AspCysPheValLysIleProArgGluProGlyAsnPro--GlyLysGlyAsnTyrTrp	80
580	QY	ATGCTCAACCCCAACAGCGAGTACACCTTCGCGACGGGGTCTTCGCGCGCCGCCCAAG	639
81	Db	ThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerPheLeuArgAGAGLys	100
640	QY	CGCTCAGCCACCCG	654
101	Db	ArgPheLysArgGln	105

Search completed: March 29, 2004, 12:05:37
Job time : 29.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 29, 2004, 12:03:47 ; Search time 59 Seconds
(without alignments)
10723.699 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189

Sequence: 1 atgaagtggagggtgttgcgt.....tggagagcgtcctagcttga 1209

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 2130338

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlsp

-MODEL=frame_n2p_model -DEV=xlsp
-Q=/cn2_1/USPTO_spool_p/US10087080/runat_29032004_114529_6667/app_query.fasta_1.1351
-DB=Published Applications AA -QWrt=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10087080 @CGN 1.13 @runat_29032004_114529_6667
-NCPU=6 -ICPU=3 -NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-FLOGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

- 1: /cn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cn2_6/ptodata/1/pubpaa/PTC_NEW_PUB.pep.*
- 3: /cn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 6: /cn2_6/ptodata/1/pubpaa/PTCUS_PUBCOMB.pep.*
- 7: /cn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

Alignment Scores:				
Pred. No.:	3.14e-92	Length:	402	
Score:	2123.00	Matches:	402	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	96.98%	Indels:	0	

ALIGNMENTS

RESULT 1
US-10-229-345-18
; Sequence 18, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CHRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-345-18

1	2123	97.0	402	12	US-10-229-345-18	Sequence 18, Appl
2	2123	97.0	402	12	US-10-274-177-18	Sequence 18, Appl
3	2123	97.0	402	15	US-10-087-080-32	Sequence 32, Appl
4	517	23.6	465	14	US-10-205-823-136	Sequence 136, Appl
5	429.5	19.6	553	10	US-09-293-862-2	Sequence 2, Appl1
6	415.5	19.0	376	14	US-10-007-280A-221	Sequence 221, Appl
7	388.5	17.7	19662	15	US-10-084-846A-6	Sequence 6, Appl1
8	387	17.7	473	9	US-09-844-353A-99	Sequence 99, Appl
9	387	17.7	473	14	US-10-177-293-212	Sequence 212, Appl
10	384.5	17.6	501	9	US-09-963-285-2	Sequence 2, Appl1
11	384.5	17.6	501	9	US-09-963-285-10	Sequence 10, Appl
12	384	17.5	494	9	US-09-963-285-7	Sequence 7, Appl1
13	372.5	17.0	19723	15	US-10-084-846A-5	Sequence 5, Appl1
14	369	16.4	19662	15	US-10-084-846A-6	Sequence 6, Appl1
15	366.5	16.7	237	14	US-10-029-386-32327	Sequence 32327, A
16	364	16.6	19652	15	US-10-084-846A-7	Sequence 7, Appl1
17	354	16.2	19608	15	US-10-084-846A-8	Sequence 8, Appl1
18	351	15.6	19652	15	US-10-084-846A-9	Sequence 9, Appl1
19	339.5	15.5	19695	15	US-10-084-846A-3	Sequence 3, Appl1
20	338.5	15.5	19725	15	US-10-084-846A-4	Sequence 4, Appl1
21	336	15.3	347	9	US-09-844-353A-100	Sequence 100, Appl
22	335	15.0	19608	15	US-10-084-846A-8	Sequence 8, Appl1
23	329	15.0	544	13	US-10-044-443-14	Sequence 14, Appl
24	323.5	14.4	19723	15	US-10-084-846A-5	Sequence 5, Appl1
25	315.5	14.0	19695	15	US-10-084-846A-3	Sequence 3, Appl1
26	306.5	14.0	417	9	US-09-963-285-4	Sequence 4, Appl1
27	305.5	14.0	534	13	US-10-044-443-11	Sequence 11, Appl
28	305	13.9	503	12	US-10-425-114-64093	Sequence 4, Appl1
29	304.5	13.6	19725	15	US-10-084-846A-4	Sequence 4, Appl1
30	302.5	13.8	777	14	US-10-156-761-10907	Sequence 10907, A
31	301.5	13.4	595	9	US-09-738-973-187	Sequence 187, App
32	301.5	13.4	595	9	US-09-854-133-187	Sequence 187, App
33	301.5	13.4	595	14	US-10-144-649A-187	Sequence 57324, A
34	294	13.4	316	12	US-10-425-114-57324	Sequence 52, Appl
35	291	13.0	641	14	US-10-138-098-52	Sequence 36798, A
36	291	13.0	641	14	US-10-294-804-4	Sequence 5, Appl1
37	288	13.2	312	12	US-10-425-114-36798	Sequence 56601, A
38	283	12.9	960	15	US-10-425-114-56601	Sequence 57887, A
39	282.5	12.9	336	12	US-10-425-114-57887	Sequence 14622, A
40	282	12.9	287	12	US-10-425-114-57887	Sequence 8, Appl1
41	281.5	12.8	375	14	US-10-156-761-14622	Sequence 47824, A
42	279.5	12.8	1259	14	US-10-260-715-8	Sequence 72062, A
43	278.5	12.7	389	12	US-10-425-114-47824	Sequence 4, Appl1
44	278.5	12.7	389	12	US-10-425-114-72062	
45	277	12.7	720	15	US-10-342-331-4	

421 CTGGCGGAGTCAACGAGTACCTCATGGGCAAGTCCCTTTTCCGGCGAGCTACAG 480
141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
481 GGCTGGCGCAACTCGTGGCGCCCAACCTTTTCGTCACAGCTCTTCGTCACAGTCTG 540
161 GlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
541 CGGACCCCTCGCGCCCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
181 ArgAspProSerArgProTrpGlyLysAsnTyrTrpMetLeuAsnProAsnSerGlu 200
601 TACACCTTCGCGCGAGCGGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
201 TyrThrPheAlaAspGlyValPheArgArgArgGlyLysArgLeuSerHisArgAlaPro 220
661 GTCCCG 720
221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240
721 GCG 780
241 AlaProAlaAlaProAlaSerProArgMetArgSerProAlaAlaGlnGluArgAla 260
781 AGCCCGCGCGCGCAATCTTCAGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
261 SerProAlaGlyLysPheSerSerSerPheAlaIleAspSerIleLeuArgLysProPhe 280
841 CGCAGCGCTCGCGCTCAGGAGCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
281 ArgSerArgArgLeuArgAspThrAlaProGlyThrThrLeuGlnTrpGlyAlaAlaPro 300
901 TGCCCG 960
301 CysProProLeuProAlaPheProAlaLeuLeuLeuProAlaAlaProCysArgAlaLeuLeu 320
961 CGGCTCTCGCGCTCAGGCG 1020
321 ProLeuCysAlaTyrGlyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGluVal 340
1021 CCACGAGCG 1080
341 ProProThrAlaProProLeuLeuLeuAlaProLeuProAlaAlaAlaProAlaLysPro 360
1081 CTCGAGCG 1140
361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrCysProLeuArgLeuProAlaAla 380
1141 CTGCGAGCGCGCTTAGTCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrProValGluThrLeu 400
1201 CTAGCT 1206
401 LeuAla 402

RESULT 3
US-10-087-080-32
; Sequence 32, Application US/10087080
; Publication NO. US20030235820A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Markowitz, Sanford David
; APPLICANT: Eos Biotechnology, Inc.
; APPLICANT: Case Western Reserve University
; TITLE OF INVENTION: No. US20030235820A1 Methods of Diagnosis of Metastatic Colorectal
; TITLE OF INVENTION: Cancer, Compositions and Methods of Screening for
; TITLE OF INVENTION: Modulators of Metastatic Colorectal Cancer
; FILE REFERENCE: 018501-000840US
; CURRENT APPLICATION NUMBER: US/10/087,080
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/272,206
; PRIOR FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: US 60/281,149
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: US 60/284,555
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 32
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: winged helix/forkhead transcription factor (HFH1)
US-10-087-080-32

Alignment Scores:
Pred. No.: 3,14e-92 Length: 402
Score: 2123.00 Matches: 402
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.98% Indels: 0
DB: 15 Gaps: 0

US-10-087-080-31 (1-1209) x US-10-087-080-32 (1-402)

QY 1 ATGAAGTTGGAGGTGTCTCTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
Db 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
QY 61 GAGGCG 120
Db 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
QY 121 GGCTCAGATGGGAGTCTCG 180
Db 41 GlySerAspGlyAspCysAlaAlaLysProSerAlaGlyGlyGlyAlaAlaArgSerThrGln 60
QY 181 GCGGAGCGGCAACAGAGTCTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 61 GlyAspGlyGluGlnSerAlaGlyGlyGlyProGlyAlaGluGluAlaIleProAlaAla 80
QY 241 GCTGCTCAGCGTGTGTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 81 AlaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaAlaGlyProGlyAlaGly 100
QY 301 GCGCGCGGAGCG 360
Db 101 GlyAlaGlySerGlyGluGlyAlaArgSerLysProTyrThrArgArgProLysProPro 120
QY 361 TACTCGTACATCCGCTCATCGCCATCGCGACTCGCGACTCGCGCGCGCGCGCGCGCGCGCG 420
Db 121 TyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeuThr 140
QY 421 CTGGCGGAGATCAACGAGTACCTCATGGGCAAGTTCCTTTTCCGGCGAGCTACAG 480
Db 141 LeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArgGlySerTyrThr 160
QY 481 GGCTGGCGCAACTCCGTGGCGCCCAACCTTTTCGTCACAGCTCTTCGTCACAGTCTG 540
Db 161 GlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPheValLysValLeu 180
QY 541 CGGACCCCTCGCGCCCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
Db 181 ArgAspProSerArgProTrpGlyLysAsnTyrTrpMetLeuAsnProAsnSerGlu 200
QY 601 TACACCTTCGCGCGAGCGGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 201 TyrThrPheAlaAspGlyValPheArgArgArgGlyLysArgLeuSerHisArgAlaPro 220
QY 661 GTCCCG 720
Db 221 ValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaProProPro 240
QY 721 GCG 780

```

Db      241 AlaProAlaProAlaSerProAsgMetArgSerProAlaArgGlnGluArgAla 260
Qy      781 AGCCCGCGGGGAACTTCCAGCTCCTTCGCCATCGACAGATCTCGCGAAGCCCTTC 840
Db      261 SerProAlaGlyLysPheSerSerPheAlaileAspSerileLeuArgLysProPhe 280
Qy      841 CGCAGCGCTCGCTCAGGGACAGCGCCCGCGGACGACCTTCAGTGGGGCGCCGCGCC 900
Db      281 ArgSerArgLysLeuArgSerThrAlaProGlyThrLeuGlnTrpGlyAlaAlaPro 300
Qy      901 TGCCCGCGGTGCGCGGTTCCTCCGCGCTCCTCCCGCGCGCGCTTCGAGGCCCTGCTG 960
Db      301 CysProProleuProAlaPheProAlaLeuLeuProAlaProCysArgAlaLeuLeu 320
Qy      961 CCCTCTGCGCGTACGGCGCGGCGAGCGCGCGCTGCGCGCGCGCGAGGCGAGGTT 1020
Db      321 ProLeuCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaVal 340
Qy      1021 CCACCGACCGCGCGCCCTCTGCTTGCACCTCTCCCGCGCGCGCCCGCCGCGAGCCA 1080
Db      341 ProProThrAlaProProleuLeuLeuAlaProLeuProAlaAlaProAlaLysPro 360
Qy      1081 CTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db      361 LeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyrcysProleuArgLeuProAlaAla 380
Qy      1141 CTCGAGCGCGCTTAGTCCGNCCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
Db      381 LeuGlnAlaAlaLeuValArgArgProGlyProHisLeuSerTyrcysProValGluThrLeu 400
Qy      1201 CTAGCT 1206
Db      401 LeuAla 402

RESULT 4
US-10-205-823-136
; Sequence 136 Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR FILING DATE: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

US-10-205-823-136
Alignment Scores:
Pred. No.: 6,37e-17 Length: 465
Score: 517.00 Matches: 157
Percent Similarity: 44.21% Conservative: 34
Best Local Similarity: 36.34% Mismatches: 132
Query Match: 23.62% Indels: 109
DB: 14 Gaps: 16

US-10-087-080-31 (1-1209) x US-10-205-823-136 (1-465)
Qy      60 GGAGGCGCGCGCGCGAGCGAGCGCGCTCCCGCTTCGGCGCGGAGAGCGACTCCCT 119
Db      39 GlyGlyGlyGlyGlyProArgLeuAlaValPro----- 49
Qy      120 GGGCTCAGATGGGACTTCGCGCGCAAGCGCTTCGCGCGGCGCGCGCGCGAGATACGCA 179
Db      50 -----AlaGlnArgArgArgA--GArgSerTyrAla 60
Qy      180 GGGCGACGGCGAACAGAG-----TGCGGG 203
Db      61 GlyGlnAspGluLeuGluAspLeuGluGluGluAspAspAspAspAspAspAspAsp 80
Qy      204 AGGCGGCGCGCGCGCGAGGAGCGATCCCGGCGAGCGAGCTCTCGACGCGGTGTGGCGGA 263
Db      81 ProProAlaGlyGlySerProAlaProProGlyProAlaProAlaAlaGlyAlaGlyAla 100
Qy      264 GGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323
Db      101 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 119
Qy      324 ACGCAGCAAGCCATATACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
Db      119 aLysAsn---ProLeuVal-----LysProProTyrcysTyrTyrTyrTyrTyrTyr 135
Qy      384 CATGGCCATCCGCGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
Db      135 rMetAlaileLeuGlnSerProLysLysArgLeuThrLeuSerGluileCysGluPheIl 155
Qy      444 CATGGCGAAGTTCCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503
Db      155 eserGlyArgPheProTyrcysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 175
Qy      504 CAACCTTTCGCTCAACGACTCTTCGTCAGGTCTCGCGACCGCGCGCGCGCGCGCGCGCG 563
Db      175 aAnLeuSerLeuAsnAspCysPheValLysileProArgGluProGlyAsnPro--Gl 194
Qy      564 CAAGGCAACTACTGGATGCTCAACCCCAACGCGAGTACACCTTCGCGCGCGCGCGCGCT 623
Db      194 yLysGlyAsnTyrcysTrpThrLeuAspProGluSerAlaAspMetPheAspAsnGlySerPh 214
Qy      624 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
Db      214 eLeuArgArgArgLysArgPheLys---ArgGlnProLeuLeuProProAsnAlaAla 233
Qy      684 CGAGGAGGCC-----CCGGCGCTCCCGCGCG 710
Db      233 aAlaGluSerLeuLeuLeuArgGlyAlaGlyAlaAlaGlyGlyAlaGlyAspProAlaAl 253
Qy      711 CCGG-----CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
Db      253 aAlaAlaAlaLeuPheProProAla--ProProProProProHisAlaTyrcysGlyP 273
Qy      743 -----CCGGCATGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
Db      273 roTyrcysGlyTyrcysGlyLeuGlnLeuProProTyrcysAlaProProSerAlaLeuPheA 293
Qy      774 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
Db      293 laAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 308
Qy      834 GCCCTTCGCGAGCGCTCGCTTCAGGAGACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 893

```

308	Db		erPro-ProProProProPro-----HisGly-----	317
894	QY		CGCGCGTCCCGCGCGTCCCGGTTCCCGCGGCT-----CCT	932
318	Db		-----AlaAlaAlaGluLeuAlaArgThrAlaPheGlyTyrA-gProHisPro	333
933	QY		CCCGCGGCGCCCTGTCAGAGGCGCTGTCGCGCTCTGCGCTACGCGCGCGGCGAGCCGCG	992
334	Db		LeuGlyAlaAlaLeuProGlyProLeuProAlaSerAlaAlaAlaGlyProGly	353
993	QY		CGCGCTGGGCGCGCGAGGC-----CGAGGTGCCACCGAC	1028
354	Db		AlaSerAlaLeuAlaArgSerProPheSerIleGluSerIleIleGlySerLeuGly	373
1029	QY		CGCGCGCCGCTCTGCTTGCCTTCCTCCCGCGCGCGCGCCCGCCCAAGCCTCCGAGG	1088
374	Db		ProAlaAlaAlaAlaAlaAlaAlaGlnAlaAlaAlaAlaAlaAlaAlaSerProSer	393
1089	QY		CCCGCGGCGCGCGCGCGCAGCTGTACTGTCGCCCTCGCGGTGCCCGCAGCCCTGCGAGGC	1148
394	Db		Pro-----SerProValAlaAlaProAlaProAlaProGlySerSerGlyGly	408
1149	QY		GGCCTTAGTCCGNGCTCCTGGCGCCGACCT	1178
409	Db		GlycysAlaAlaGlnAlaAlaValGlyPro	418

RESULT 5
 US-09-292-862-2
 ; Sequence 2, Application US/09292862
 Publication No. US20030013082A1

```

RESULT 5
US-09-292-862-2
; Sequence 2, Application US/09292862
; Publication No. US20030013087A1
; GENERAL INFORMATION:
; APPLICANT: Walter, Michael A.
; APPLICANT: Jordan, Tim
; APPLICANT: Raymond, Vincent
; TITLE OF INVENTION: NOVEL MUTATIONS IN THE PRCAC3 GENE FOR
; TITLE OF INVENTION: DIAGNOSIS AND PROGNOSIS OF GLAUCOMA AND ANTERIOR SEGMENT
; TITLE OF INVENTION: DYSGENESIS

```

```

, FILE REFERENCE: 07540/020003
, CURRENT APPLICATION NUMBER: US/09/292,862
, CURRENT FILING DATE: 1999-04-16
, EARLIER APPLICATION NUMBER: 60/084,784
, EARLIER FILING DATE: 1998-05-08
, EARLIER APPLICATION NUMBER: 60/082,206
, EARLIER FILING DATE: 1998-04-17
, NUMBER OF SEQ ID NOS: 6
, SOFTWARE: FastSEQ for Windows Version 3.0
, SEQ ID NO 2
, LENGTH: 553
, TYPE: prt
, ORGANISM: Homo sapiens
, US-09-292-862-2

```

Alignment Scores:	
Pred. No.:	7.87e-13
Score:	429.50
Length:	553
Matches:	146
Percent Similarity:	38
Best Local Similarity:	41.82%
Mismatches:	124
Query Match:	132
Indels:	19.62%
Gaps:	18
DB:	10

[illegible]

Db	58	GlyGlyMetAlaArgAlaTyrgly	-----ProTyThrProGlnProGlnPro	73
Qy	352	-----AAGCCCCCTACTCGTACATCGCGCTCATCGCCATCGGCATCGCGGAC	399	
Db	74	LysAspMetValIysProProTySerTyrlleAlaLeuIleThrMetAlaIleGlnAsn	93	
Qy	400	TGGCGGGCGGGCGCTTGACGCTGGCGGAGATCAACGAGTACTCATGGCAAGTATCCCC	459	
Db	94	AlaProAspIlyAlaIleThrLeuAsnGlyIleTyGlnPheIleMetAspArgPhePro	113	
Qy	460	TTTTTCCGGCGCATACAGGGCTCGGCAACTTCGTGCGGCCACAACTTTCGTCTCAAC	519	
Db	114	PheTyArgAspAsnIysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn	133	
Qy	520	GACTGCTTCGTCAAGTGCTGGCGACCCCTCGCGCCCTGGGGCAAGGACAACTACTGG	579	
Db	134	GluCysPheValIysValProArgAspAspLysIlePro---GlyIysGlySerTyTrp	152	
Qy	580	ATGCTCAACCCCAACAGCGAGTACACTTTGCGCGAGCGGGTCTTTCGCGCCCGCGCGCAAG	639	
Db	153	ThrLeuAspProAspSerTyAenMetPheGluAsnGlySerPheLeuArgArgArgArg	172	
Qy	640	CGCTCAGC-----CACCGCGCG	657	
Db	173	ArgPheLysLysAspAlaLeuLysAspLysGluLysAspArgLeuHisLeuLys	192	
Qy	658	CCGCTCCCGCGCGCGCGCTCGCGCCGAGAGCGCCCGCGCTCCCGCGC-----	708	
Db	193	GluProProProGlyAlaSerProArgProAlaProGluGlnAlaAspGlyAsn	212	
Qy	709	GCCTCGCGCGCGCGCGCGCGC-----CGGCTCGCGCGCG	747	
Db	213	AlaPro-GlyProGlnProProProValArgIleGlnAspIleLysThrGluAsnGlyTh	232	
Qy	748	ATCGCTCGCGCGCGCGCGAGGAGC---GGCCAGCGCCCGGCAAGTCTCCAGC	804	
Db	232	rCysProSerProProGlnProLeuSerProAlaAlaLeuGlySerGlySerAlaAl	252	
Qy	805	TCCTTCGCATCGACAGCATCTCGCAGACCTTCGCGACCGCTCGCTCAGGACACG	864	
Db	252	aAlaValProLysIleGluSerProAspSerSerSerSerLeuSerGlySerSe	272	
Qy	865	GCCCGCGGGA-----CGAGCTTCAGT-----GGGCGCGC	894	
Db	272	rProGlySerLeuProSerAlaArgProLeuSerLeuAspGlyAlaAspSerAlaPr	292	
Qy	895	GCGCCTTCGCGCGCGCTGCGCGCTTCGCGCGCTCTCC-----	935	
Db	292	oProPro-ProAla-----ProSerAlaProProProHisHisSerGlnGlyP	308	
Qy	936	-----CGCGCGCCCTGCAGGGCCCTGCTGCGCTC	966	
Db	308	heSerValAspAenIleMetThrSerLeuArgGlySerProGlnSerAlaAlaGluL	328	
Qy	967	TGCGCTGATCGCGCGGCGAGCGCGCGCTGGCGCGCGAGCGCGAGGTCACACG	1026	
Db	328	eu-----SerSerGlyLeuLeuAlaAlaAlaSerS	340	
Qy	1027	ACGCGCGC-----GCCCTCTCT-----	1043	
Db	340	erArgAlaGlyIleAlaProProLeuAlaLeuGlyAlaTySerProGlyGlnSerSerL	360	
Qy	1044	-----GCTTGACCTCTCCCGCGCGCGCGCGCGC	1074	
Db	360	euTySerSerProCysSerGlnThrSerSerAlaGlySerSerGlyGlyGlyGlyG	380	
Qy	1075	AAGCCACTCGAGGCCCGCGCGCGCGCGGACCTGTAC---TGCCCCCTGCGG---	1128	
Db	380	lyAlaGlyAlaAlaGlyGly-AlaGlyGlyAlaGlyThrTyHisCysAsnLeuGlnAla	399	
Qy	1129	-----CTGCGCGCAGCGCTG	1143	
Db	400	MeSerLeuTyAlaAlaGlyGluArgGlyGlyHisLeuGlnGlyAlaProGlyGlyAla	419	

Qy 1144 CAGCGCGCTTAGTCCGNCCTCTCGGCCCGCAGCACTGTCGTACCGCGTGGAGAGC 1197
Db 420 GlycylserAlaValAspProLeuProAspTyrSerLeuProValThr 437

RESULT 6
US-10-007-280A-221
; Sequence 221, Application US/10007280A
; Publication No. US20030059784A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Hervé
; APPLICANT: Salceda, Susana
; APPLICANT: Chenghua, Liu
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0257
; CURRENT APPLICATION NUMBER: US/10/007,280A
; CURRENT FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 60/246,640
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 221
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-007-280A-221

Alignment Scores:
Pred. No.: 3 78e-12 Length: 376
Score: 415.50 Matches: 128
Percent Similarity: 42.59% Conservative: 33
Best Local Similarity: 33.86% Mismatches: 118
Query Match: 18.98% Indels: 99
DB: 14 Gaps: 15

US-10-087-080-31 (1-1209) x US-10-007-280A-221 (1-376)

Qy 232 CGCGCAGCACTGCTGCGAGCGTGGTGGCG-----GAGGGCGCG 270
Db 8 ProgluAspAlaAlaGlyAlaLeuLeuAlaProgluThrGlyArgThrValLysGluPro 27
Qy 271 GAGCGCGGGCGGGCGGGCCAGCGCGGGCGGGCGGGAGCGCGAGGTGGTGGCGAGC 330
Db 28 GluGlyProProProSerProGlyLysGlyGlyGlyGlyGlyGlyGlyGlyThrAlaProGlu 47
Qy 331 AAGCATATACGGCGGGCGGCCAAGCCCCCTACTCGTACATCGGCTCATCGGCATCGGCATGGCC 390
Db 48 LysProAspProAlaGln--LysProProTyrSerTyrValAlaLeuAlaMetAla 66
Qy 391 ATCCGCACATCGGGCGGGCGGGCGGCTTCACGCTGGCGGAGATCAACGAGTACCTCATGGGC 450
Db 67 IleArgGluSerAlaGluLysArgLeuThrLeuSerGlyIleTyrGlnTyrIleAla 86
Qy 451 AAGTTCCTCTTTTCGCGGAGCTACACGGGCTGGCGCACTCGGTGGCCCAACACCTT 510
Db 87 LysPheProPheTyrGluLysAsnLysGlyTyrGlnAsnSerIleArgHisAsnLeu 106
Qy 511 TCGCTCAACGACTGCTTCGTCAAGGTGCTGGCGACCCCTCGCGGCCCTCGGGCAAGGAC 570
Db 107 SerLeuAsnGluCysPheIleLysValProArgGlnGlyGlyGlu---ArgLysGly 125
Qy 571 AACTACTGGATGTTCAACCCCAACAGCGAGTACACCTTCGCCGACGGGTGTTCCCGCCG 630
Db 126 AsnTyrTrpThrLeuAspProAlaCysGluAspMetPheGluLysGlyAsnTyrArgArg 145
Qy 631 CGCGCAAGCGCCTCAGCCACCGCGCGCGTCCCG----- 666
Db 146 ArgArgArgMetLysArgProPheArgProProProAlaHisPheGlnProGlyLysGly 165
Qy 667 -----GCCTCCGGCTGGCGCCGAGGAGGCGCGCTCC-- 703
Db 166 LeuPheGluValArgIleGluValAlaAla-GluGlyCysGluValAlaGluValAlaAspGlu 185

Qy	704	-----CCGCGCGCCGCGC-----	715
		:::	
Db	185	yTyGlyTyLeuAlaProProLyStyTyLeuGlnSerGlyPheUenAenSerTrpPr	205
Qy	716	-CGCGCGCGCGCGCG--CCCGGCTCGCCCGCATGCGTGGCTCGCGCGCGCGCGAG	771
Db	205	oLeuProGlnProProSerProMetProTyTyAlaSerCysGlnMetAlaAlaAla	225
		:::	
Qy	772	GAGCGCGCAGCGCGCGGCGAGTTCCTCAGCTCTTCGCCATCGACAGCATCC	826
		:::	
Db	225	aAlaAlaAlaAlaAlaAlaAlaAlaGlyProGlySerProGlyAlaAlaAlaVal	245
Qy	827	-----TGCGCAAGCCCTCCGAGCGCTCGCTCAGGGACACGG-----	865
Db	245	lLysGlyTyLeuAlaGlyProAlaAlaSerTyGlyProTyThrArgValGlnSerMetAl	265
Qy	866	-----CCCCGGGA-----CGAGCTTCAGTGGGGCGCGCGCTCGCCGCGCTGCC	915
		:::	
Db	265	aLeuProProGlyValValAlaSerTyAenGlyLeuGlyGlyProProAlaAlaProPr	285
Qy	916	GCGTTCGCGCGCTCTCCCGCGCGCCCTGCAGGGCCCTGCTGCCGCTCTGCCGCTAC	975
Db	285	oProProProHisProHisPro-----	292
Qy	976	GGCGCGCGGAGCGCGCGCGCTGGGCGCGCGGAGCGAGTGCACCGACCCGCGCG	1035
		:::	
Db	293	-----HisProHisAlaHisHis-----	298
Qy	1036	CCCTCTCTGCTTGCACTCTCCCGGGCGCGCCCGCCAGGCACATCCGAGGCGCGCG	1095
		:::	
Db	299	-----LeuHis---AlaAlaAlaAlaProProProAlaProHisGly-	313
Qy	1096	GCCGCGCGCGCAGCACTGTACTGCCCTCGCGTGCCTCGCGAGCCCTGCAGGCGCGCTTA	1155
Db	314	-----AlaAlaAlaProProProGlyGlnLeuS	323
Qy	1156	GTCGCGCTCTGGCGCGCACCTCTGTCACCGGTGGAGAGCGTCTCTAGC	1205
		:::	
Db	323	erProAlaSer---ProAlaThrAlaAlaProProAlaProAlaProThr	338
RESULT 7			
US-10-084-846A-6			
; Sequence 6, Application US/10084846A			
; Publication No. US2004006026A1			
; GENERAL INFORMATION:			
; APPLICANT: WEITNAUER, GABRIELE			
; APPLICANT: MUHLENWEG, AGNES			
; APPLICANT: TREFFZER, AXEL			
; APPLICANT: BECHTHOLD, ANDREAS			
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES			
; FILE REFERENCE: 1974-005			
; CURRENT APPLICATION NUMBER: US/10/084,846A			
; CURRENT FILING DATE: 2003-02-25			
; PRIOR APPLICATION NUMBER: PCT/EP01/09815			
; PRIOR FILING DATE: 2001-08-24			
; PRIOR APPLICATION NUMBER: DE 101 09 166.4			
; PRIOR FILING DATE: 2001-02-25			
; NUMBER OF SEQ ID NOS: 120			
; SOFTWARE: PatentIn Ver. 3.2			
; SEQ ID NO 6			
; LENGTH: 19662			
; TYPE: PRT			
; ORGANISM: Streptomyces viridochromogenes			
; FEATURE:			
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand			
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.			
US-10-084-846A-6			
Alignment Scores:			
Pred. No.: 3,84e-11 Length: 19662			
Score: 388.50 Matches: 165			
Percent Similarity: 39.41% Conservative: 21			

```

; FEATURE:
; OTHER INFORMATION: protein 1: amino acid sequence encoded by coding strand 2
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6

Alignment Scores:
Pred. No.: 3,848-11 Length: 19682
Score: 388.50 Matches: 165
Percent Similarity: 39.41% Conservative: 21

```

Best Local Similarity: 34.96% Mismatches: 162
Query Match: 17.75% Indels: 124
DB: 15 Gaps: 28

US-10-087-080-31 (1-1209) x US-10-084-846A-6 (1-19662)

```
QY 18 CGTCTCCGCGCGCCCAAGCGGACAGAGGCGAGTGAAGCTGAGGCGCGCGG 71
DB 12919 ArgProGlyArgSerProArgGlyArgGlyCys---ProGlyThrGlyGlyCysArg 12937
QY 72 -----CGGACAGCGCGCGCTCCCG 92
DB 12938 ArgProAlaCysCysAlaThrAlaAlaGlyArgCysProArgArgArgAlaCysPro 12957
QY 93 GCTGTCCGCGCGGAGACGA-----CTCCCTGGGCTCAGATGG--- 131
DB 12958 AlaAlaCysGlySerArgArgGlyArgAlaArgProLeuAlaProArgAArgTrpAsn 12977
QY 132 GGAAGTCCGCGCGCAAGCGTCCGCGGCGG-----CGGCGCCAGAGATACGA 179
DB 12978 AlaThrAlaGlyGlyAlaAlaArgGlyArgAlaAlaGlyLeuArgGlySerArgSerAla 12997
QY 180 GGG-----CGACGCGCAACAGAGTCCGCGGAGCGCGCGCGCGCGGAGGA 224
DB 12998 GlyAlaProThrSerProArgSerArgGlySerGlyArgArgProArgArgGly 13017
QY 225 -----GCGATCCGCGCGAGCGTCTGACGCGTGTGCGCGAGGCGCGG----- 271
DB 13018 ArgSerGlyPro-ArgAlaGlyCysAlaValAlaSerVal-ArgArgThrArgProPro 13037
QY 272 -----AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 323
DB 13037 ArgValArgProAlaSerArgGlySerSerAlaArgProArgGlyAlaGly----- 13054
QY 324 ACGACGAAGGCATATACGCGCGCG-----CAAGCGCGCGCGCGCGCGCGCGCGTAT 380
DB 13055 --ThrAlaArgArgThrProGlyArgProArgProArgProArgSerArgArgAla 13074
QY 381 CGCCATGG-----CCATCCGCGACTCCGCGCGCGCGCGCGCGCGCGCGCGCG 428
DB 13074 laProGlyCysAlaGlySerProProSerThrArgArg-----ArgA 13088
QY 429 GATCAACGAGTACCTATGCGCAAGTTCCTTTTCGCGCGCGCGCGCGCGCGCGCGCG 486
DB 13088 rGSer-----ProTrpProAlaGlyPro--ArgAlaProSerCysArgAlaAla- 13103
QY 489 CAACTCCGTCGCGCGCAACCTTCGCTCAACGACTGCTTCCTCAAGTGTCTCCGCGACCC 548
DB 13104 -----CysAlaValArgTrpAspSerAlaArgArgArgProArgAlaAlaGlyArg 13121
QY 549 CTCGCGCGCGCGCGCGCAAGCAACTACTGATGCTCAACCCCA-----ACAG 596
DB 13121 laGlyGlyProCysAla-----GlyCysGlyArgProArgArgArgProA 13137
QY 597 CGAGTACACTTCGCGCGAGGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
DB 13137 laValArgProAlaProArgSerSer-----GlyA 13148
QY 657 GCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 713
DB 13148 rGArgProSerArgProArgArgGlyHisArgArgSerArgTrpThrSerSerArgThrA 13168
QY 714 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
DB 13168 rGTrpProArg---ProProArgHisCysProArgAlaAlaArgProArgArgArgA 13187
QY 774 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
DB 13187 laAspProSerProSerArgHisAlaProArgProArg----- 13199
QY 834 GCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
DB 13200 -----AlaAlaAlaAlaThrTyrGlyArgSerAlaGlyProArgArgSerArgSerAlaP 13218
```

```
QY 894 CCGCGCCTCCCGCGCGCTGCCCGCTCCCGCGCTCCCGCGCGCGCGCGCGCGCGCG 953
DB 13218 roArgArgArgArgArgArgSerHisArgProSerGlyArgArgArgAlaArgA-gp 13238
QY 954 CTGTGCTCCGCTCTCGGCTACGGCGCG---GCGAGCGCGCGCGCGCTGGCGG---CGCG 1007
DB 13238 roValThrAlaThrAlaArgArgAlaAlaSerArgArgSerTrpSerGlyArgA 13258
QY 1008 CGAGCGCGAGGTGCGCCACCGCGCGCGCGCGCGCGCTCTCTCTCCCTCCCGCGCGCG 1067
DB 13258 rGArgProArgCysGlyArgPro-----ProAlaAlaSerAlaGlyAlaArgThrArgS 13276
QY 1068 CCGCGCAACCGCTCCGAGGCGCGCGCG-----CGCG 1100
DB 13276 erProProThrProAlaProValArgArgSerProArgArgArgProAlaArgLeuProA 13296
QY 1101 CGCGCGCGCGCTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT- 1148
DB 13296 laValArgGlyProSerArgProThrGly-ThrArgAspProArgThrProArgArgAsp 13315
QY 1149 GCGCTTAGTCCGCGTCC-----TGCGCGCGCA 1175
DB 13316 GlyArgHisProAlaAlaArgProArgArgAspArgAlaArgValProGluTrpProAla 13335
QY 1176 -----CCTGCTGACCGCGT 1190
DB 13336 ArgAlaAlaArgArgProValArgProGly 13345
```

RESULT 8

US-09-844-353A-99
; Sequence 99, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruykun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowsek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351005
; CURRENT APPLICATION NUMBER: US/09/844,353A
; CURRENT FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 08/857,076
; PRIOR FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-844-353A-99

Alignment Scores:
Pred. No.: 7,93e-11 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 9 Gaps: 15

US-10-087-080-31 (1-1209) x US-09-844-353A-99 (1-473)

```
QY 70 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
DB 88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
QY 130 GGGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189
```

Db 107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnGlnAlaAla 122
Qy 190 GAACAGAGTGGCGGAGCGCGCGCGCGAGAGCGATCCCGCAGCAGCAGCTGCTGCA 249
Db 123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
Qy 250 GCGGTGGTGGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
Db 143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
Qy 310 AGCGGCGAGGTGCAGCGAGAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
Db 160 AlaLysThrPhenylsArgSerTyrProHisAla-----LysProProTyrSerTyr 176
Qy 370 ATCGCGCTATCCCATCGCATCCCGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 429
Db 177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
Qy 430 ATCAACAGAGTACCTATGGCAAGTTCCTCTTTTCGCGGAGCAGTACACGGGCTGGCG 489
Db 197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTrpGln 216
Qy 490 AACTCGGTGGCGCACACCTTCCTCAAGACTCTCTCAAGGTGCTGCGCGCACCC 549
Db 217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValysValalaArgSerPro 236
Qy 550 TCGCGCCCTGGCGGCAAGCAACTACTGATGCTCAACCCCAACAGCAGTACACCTTC 609
Db 237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
Qy 610 GCGAGGGGTCTTCG 669
Db 256 GluAsnGlyCysTyrLeuArgGlnLysArg----- 266
Qy 670 CCGCGGTGGCG 729
Db 267 -----PheLysCysGlnLysGlnProGly-----AlaGlyGly 277
Qy 730 GCGCGCGCTTCG 783
Db 278 GlyGlyGlySerGlySerGlyGlySerGlyAlaLysGlyGlyProGluSerArgLysAsp 297
Qy 784 CCGCGCGCAAGTCTTCAGCTCTTCGCGCATCGACGATCTCGCGCAAGCCCTCCCG 843
Db 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315
Qy 844 AGCGGTGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
Db 316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326
Qy 904 CCGCGGTGCGCGCGGTTCGCGCGCTCCCTCCCGCGCGCGCGCGCGCGCGCGCGCG 963
Db 327 ProGlyProAlaLysArgProGlnThrLeuAsp----- 337
Qy 964 CTCTGCGGTACG 1023
Db 338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGluLeuLysThrPro 353
Qy 1024 -----CCGACCG 1062
Db 354 AlaSerSerThrAlaProPheIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373
Qy 1063 GCG 1122
Db 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
Qy 1123 CTGCGGTGCG 1182
Db 390 -----LysGlyAspProHisTyrSer 396
1183 TAC 1185
Qy 397 Phe 397

RESULT 9
US-10-177-293-212
; Sequence 212, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Fuzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-212

Alignment Scores:
Pred. No.: 7,93e-11 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 14 Gaps: 15

US-10-087-080-31 (1-1209) x US-10-177-293-212 (1-473)
Qy 70 GCGCGCAGCAGCGCGCGCTCCCGCTGTCGCGCGCGGAGCAGCAGCTCCCTGGGCTCAGAT 129
Db 88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
Qy 130 GGGGACTGCGCGGCAAGCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189
Db 107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnGlnAlaAla 122
Qy 190 GAACAGAGTGGCGGAGCG 249
Db 123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
Qy 250 GCGGTGGTGGCGGAGCG 309


```
Db 143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
Qy 310 ACCGCGAGGTGACGACGAGCATATACGCGCGGCCACAGCCCTCTCTCTCTAC 369
Db 160 AlaLysThrPheLysArgSerTyrProHisAla-----LysProProTyrSerTyr 176
Qy 370 ATCGCGCTCATCGCATCGCATCGCATCGCGCGCGCGCGCGCGCGCGCGCGCG 429
Db 177 HisSerLeuLeuThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
Qy 430 ATCAAGAGTACTCATGCGGCAAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 489
Db 197 IleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGlnAsnGlnInArgTrpGln 216
Qy 490 AACTCGCTCGCGCCACCACTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 549
Db 217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
Qy 550 TCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 609
Db 237 AspLysPro---GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
Qy 610 GCCGAGGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Db 256 GluAsnGlyCysTyrLeuArgArgGlnLysArg----- 266
Qy 670 CCGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Db 267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
Qy 730 GCCCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
Db 278 GlyGlyLysSerGlySerGlyAlaLysGlyLysGlyProGluSerArgLysAsp 297
Qy 784 CCGCGCGCGCAAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 843
Db 298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgLysValHis 315
Qy 844 AGCGTGCCTCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
Db 316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326
Qy 904 CCGCGGCTCGCGCGGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 963
Db 327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
Qy 964 CTCTCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1023
Db 338 -----HisSerGlyAlaThrAlaThrGlyLysAlaSerGluLeuLysThrPro 353
Qy 1024 -----CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
Db 354 AlaSerSerThrAlaProProIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373
Qy 1063 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
Db 374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
Qy 1123 CTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182
Db 390 -----LysGlyAspProHisTyrSer 396
Qy 1183 TAC 1185
Db 397 Phe 397
```

RESULT 10

```
US-09-963-285-2
; Sequence 2, Application US/09963285
; Patent No. US20020090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbeck, Sven
; APPLICANT: Krook, Katarina
```

```
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-285-2

Alignment Scores:
Pred. No.: 1,03e-10 Length: 501
Score: 384.50 Matches: 127
Percent Similarity: 45.86% Conservative: 39
Best Local Similarity: 35.08% Mismatches: 118
Query Match: 17.57% Indels: 78
DB: 9 Gaps: 15

US-10-087-080-31 (1-1209) x US-09-963-285-2 (1-501)
Qy 259 GCGAGGAGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
Db 27 AlaAlaGlySerTyrGlyGlyMetAlaSerProMetGlyValTyrSerGlyHisProGlu 46
Qy 310 -----AGCGCGAGGGTGCACGAGCAAGCCATATACGCGCGCGCGCGCG 351
Db 47 GlnTyrSerAlaGlyMetGlyArgSerTyrAlaProTyrHisHisGlnProAlaAla 66
Qy 352 -----AAGCCCCCTACTCTGATACGCGCTCATCGCGCTCATCGCGCATCGCG 396
Db 67 ProLysAspLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGln 86
Qy 397 GACTCGCGCGCGCGCGCTTGACGCTGCGGAGATCAACAGTACTCATCGGCAAGTTC 456
Db 87 AsnAlaProGluLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhe 106
Qy 457 CCCTTTTTCGCGCGGCTACACGCGCTGCGCAACTCCGTGCGCGCACACCTTTTTCGCTC 516
Db 107 ProPheTyrArgGluAsnLysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeu 126
Qy 517 AACGACTGCTTCGTCAGGTGCTGCGCGACCCCTCGCGCGCGCGCGCGCGCGCG 576
Db 127 AsnGluCysPheValLysValProArgAspAspLysLysPro---GlyLysGlySerTyr 145
Qy 577 TGGATGTCACCCCAACACGAGTACACCTTCGCGAGCGGGTCTTCGCGCGCGCGCG 636
Db 146 TrpThrLeuAspProAspSerTyrAsnMetPheGluAsnGlySerPheLeuArgArg 165
Qy 637 AAGCGCTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 696
Db 166 ArgArgPheLysLysLys-----AspValSerLysGluLysGluArgAla 181
Qy 697 GGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
Db 182 HisLeuLysGluProProProAlaAlaSerLysGlyAlaProAlaThrProHisLeuAla 201
Qy 754 TCGCGCGCGCGCGCGAGGAGCGC-----GCCAGCCCGCGCG 789
Db 202 AspAlaProLysGluAlaGluLysValLysValLysSerGluAlaAlaSerProAla 221
Qy 790 -----GCCAAGTTCCTCAGCTCCTTCGCGCATCGACAGCA----- 823
Db 222 LeuProValIleThrLysValGluThrLeu-SerProGluSerAlaLeuGlnGlySerPr 241
```

QY 824 ----TCTGCGAAGCCCTTCCGACGCGTGCCTCAGGACACGGCCCGCGGACGACG 879
 Db 241 oArgSerAlaAlaSer-ThrProAlaGlySerProAspGlySerLeuProGluHisA 261
 QY 880 CTTCACTGGGCGCGCGCTCCGCGCGCTGCC-----CGCGTTCCCG 924
 Db 261 laAla---AlaProAnGlyLeuProGlyPheSerValGluAnlleMetThrLeuArgT 280
 QY 925 GCCTCTCTCCCGCGCGCTGCAGGCGCTGCTCCGCTCTGCGGTACGCGCGCGGC 984
 Db 280 hrSerProGlyGlyGluLeu-----287
 QY 985 GAGCGCGCGCGTGGCGCGCGAGCGGAGTCCACCGACCGCGCGCGCTCTCTG 1044
 Db 288 --SerProGlyAlaGlyArgAla-GlyLeuValProProLeuAlaLeuProTyrAla 306
 QY 1045 CTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1104
 Db 307 AlaAlaPro---ProAlaAlaTyrGlyGlnProCysAlaGlnGlyLeuGluAlaGlyAla 325
 QY 1105 GCGCACTGTAC---TGCCCGCTGCGG-----1128
 Db 326 AlaGlyGlyTyrGlnCysSerMetArgAlaMetSerLeuTyrThrGlyAlaGluArgPro 345
 QY 1129 -----CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173
 Db 346 AlaHisMetCysValProProAlaLeuAspGluAlaLeuSerAspHisProSerGlyPro 365

RESULT 11

US-09-963-285-10
 ; Sequence 10, Application US/09963285
 ; Patent No. US20020090707A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Enerbck, Sven
 ; APPLICANT: Krook, Katarina
 ; APPLICANT: Rondahl, Lena
 ; APPLICANT: Wasserman, Wyeth
 ; TITLE OF INVENTION: PROMOTER SEQUENCES
 ; FILE REFERENCE: 13425-042001
 ; CURRENT APPLICATION NUMBER: US/09/963,285
 ; CURRENT FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: SE 0004102-0
 ; PRIOR FILING DATE: 2000-11-09
 ; PRIOR APPLICATION NUMBER: US 60/238,897
 ; PRIOR FILING DATE: 2000-10-10
 ; PRIOR APPLICATION NUMBER: SE 0003435-5
 ; PRIOR FILING DATE: 2000-09-26
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 501
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-963-285-10

Alignment Scores:
 Pred. No.: 1.03e-10 Length: 501
 Score: 384.50 Matches: 127
 Percent Similarity: 45.86% Conservative: 39
 Best Local Similarity: 35.08% Mismatches: 118
 Query Match: 17.57% Indels: 78
 DB: 9 Gaps: 15

US-10-087-080-31 (1-1209) x US-09-963-285-10 (1-501)

QY 259 GCGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 309
 Db 27 AlaAlaGlySerTyrGlyGlyMetAlaSerProMetGlyValTyrSerGlyHisProGlu 46
 QY 310 -----ACGCGGAGGTGCAGCGACGACCATATACGCGCGCGCGCGCG 351
 Db 47 GlnTyrSerAlaGlyMetGlyArgSerTyrAlaProTyrHisHisGlnProAlaAla 66

RESULT 12

US-09-963-285-7
 ; Sequence 7, Application US/09963285
 ; Patent No. US20020090707A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Enerbck, Sven
 ; APPLICANT: Krook, Katarina
 ; APPLICANT: Rondahl, Lena

QY 352 -----AAGCCCGCTACTCGTACATCGCGCTCATCGCATGGCCATCGC 396
 Db 67 ProLysAspLeuValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGln 86
 QY 397 GACTCGCGCGCGCGCTTACGCTCGCGGAGATCAACAGTACCTCATCGGCAAGTTC 456
 Db 87 AsnAlaProGluLysLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhe 106
 QY 457 CCCTTTTTCGCGCGCGCTACACGGGCTGGCGCACTCCGTCGCCACACACTTTCCTC 516
 Db 107 ProPheTyrArgGluAnLysGlnGlyTyrPheGlnAsnSerIleArgHisAsnLeuSerLeu 126
 QY 517 AACCACTGCTTCGTCAGGTGCTCGCGGACCTCGCGCGCTCGGGCAAGCAACATAC 576
 Db 127 AsnGluCysPheValLysValProArgAspAspLysPro---GlyLysGlySerTyr 145
 QY 577 TGGATGCTCAACCCCAACAGAGTACACCTTTCGCCGCGCGGGTCTTCCGCGCGCGCGC 636
 Db 146 TrpThrLeuAspProAspSerTyrAsnMetPheGluAnGlySerPheLeuArgArgArg 165
 QY 637 AAGCGCTCAGCCACCGCGCGCGCTCCCGCGCGCGGTGCGCGCGCGCGCGCGCGCG 696
 Db 166 ArgArgPheLysLysLys-----AspValSerLysGluLysGluGluArgAla 181
 QY 697 GGCCTCCCG 753
 Db 182 HisLeuLysGluProProProAlaAlaSerLysGlyAlaProAlaThrProHisLeuAla 201
 QY 754 TCG 789
 Db 202 AspAlaProLysGluAlaGluLysValValLysSerGluAlaAlaSerProAla 221
 QY 790 -----GGCAAGTTCTCCAGTCTCCCATCGCTTCCCATCGACACA-----823
 Db 222 LeuProValIleThrLysValGluThrLeu-SerProGluSerAlaLeuGlnGlySerPr 241
 QY 824 ----TCCTGCGCAAGCCCTTCGCGCGCGCTCGCGCGCGCTCGCGCGCGCGCGCGCG 879
 Db 241 oArgSerAlaAlaSer-ThrProAlaGlySerProAspGlySerLeuProGluHisA 261
 QY 880 CTTCACTGGGCG 924
 Db 261 laAla---AlaProAnGlyLeuProGlyPheSerValGluAnlleMetThrLeuArgT 280
 QY 925 GCGTCTCTCCCG 984
 Db 280 hrSerProGlyGlyGluLeu-----287
 QY 985 GAGCGCGCGCGCTGGCTCT 1044
 Db 288 --SerProGlyAlaGlyArgAla-GlyLeuValProProLeuAlaLeuProTyrAla 306
 QY 1045 CTTGCACTCTCCCG 1104
 Db 307 AlaAlaPro---ProAlaAlaTyrGlyGlnProCysAlaGlnGlyLeuGluAlaGlyAla 325
 QY 1105 GCGCACTGTAC---TGCCCGCTGCGG-----1128
 Db 326 AlaGlyGlyTyrGlnCysSerMetArgAlaMetSerLeuTyrThrGlyAlaGluArgPro 345
 QY 1129 -----CTGCG 1173
 Db 346 AlaHisMetCysValProProAlaLeuAspGluAlaLeuSerAspHisProSerGlyPro 365

QY 68 CGGGCGGACGACGCGCGCGTCCCGCTGTCGCGCGGAGACGACACTCCCTGGCTCAG 127
 Db |||:|||||
 6026 ProGlyAlaSerThr-----CysTrpCysHisAlaArgArgSerProTrpAlaArg 6043
 QY |||:|||||
 128 -----ATGGGAGCTCGCGCGCAAGCGCTCGCGGGCGGGCGG 166
 Db |||:|||||
 6044 ProAlaArgTrpArgProThrAlaGlyAlaSerArgSerProArgArgProAspAlaThr 6063
 QY |||:|||||
 167 CCAGAGATACGAGGCGAGCGGAAAGAGTGCGGGAGCGGGCGG-----GCGCGG 220
 Db |||:|||||
 6064 ProLeuProArgArgProThrAlaThrAlaAlaAlaArgAlaAlaCysTrpSerSer 6083
 QY |||:|||||
 221 AGAGGCGATCCGCGACGACGCTGTCAGCGGTGGTGGCG-----AGG 265
 Db |||:|||||
 6084 SerCysArgThrProTrpAlaThrValThrGlyCysTrpArgCysCysAlaAlaAlaArg 6103
 QY |||:|||||
 266 GCGCGAGGCGGGGGCGG-----CGGGCGAGCGGGCGGGCGG 304
 Db |||:|||||
 6104 SerThrArgThrGlyAlaProThrAlaSerTrpProProAlaGlyArgProArgSerThr 6123
 QY |||:|||||
 305 CGGGAGCGGAGGGTGACGAGCA-----AGCATATACGCGC 346
 Db |||:|||||
 6124 SerGlyAlaProArgProArgAlaSerLeuProThrProLeuThrSerSerArgProThr 6143
 QY |||:|||||
 347 GGCCCAAGCCC-----CCTACTCGTACATCGCGCTCATCGCATGGCCATCC 394
 Db |||:|||||
 6144 AlaProValProAlaSerAlaThrProTrpArgSerAlaArgSerProProSerThrAsp 6163
 QY |||:|||||
 395 GCGACTCG-----CGGGCGGCGCTTGACGCTGGCGGAGATCAACGAGTACCTCATGG 448
 Db |||:|||||
 6164 ValThrAlaProValArgSerAlaArgSerGlyArgSerProThrSerGlyIleTrp 6183
 QY |||:|||||
 449 -----GCAAGTTCCCTTTTCGCGCGCAGCTACACGGGTGGC 487
 Db |||:|||||
 6184 ArgGlyProProGlySerProGlySerSerArgArgSerSerProSerThrArgThrGly 6203
 QY |||:|||||
 488 GCAACTCG---TCGCGCAACACTTTCGCTCAACGACTGCTTCGTCAGGTGCTGCGG 544
 Db |||:|||||
 6204 SerArgProAlaCysTrpThrAlaThrArg----- 6213
 QY |||:|||||
 545 ACCCTCGCGCGCTGGGCAAGGAGCAACTACTGGA-----TGCTCAACC 589
 Db |||:|||||
 6214 ThrProArgSerThrGlyProAlaSerThrSerGlyTrpArgProGlyProCysArgGly 6233
 QY |||:|||||
 590 CCAACAGCGAGTACACTTCG-----CCGACGGGGTCTCCCGCGCGCGCGCA 637
 Db |||:|||||
 6234 ProSerValArgThrProAlaGlyArgLeuCysProValSerAlaThrAlaAlaProTrp 6253
 QY |||:|||||
 638 AGCGCTCAGCCACCGCGCGGTCGCCGCGCGGCTGCGCGCGAGGAGGCGCGG 697
 Db |||:|||||
 6254 ProMetTrpSerSerAsnArgArgProProArgProPro---AlaProArgArgSerArg 6272
 QY |||:|||||
 698 GCCTCCCGCGCGCGCGCGCGCGCGCGCG-----CCCGCGCT 739
 Db |||:|||||
 6273 ProGluProCysSerArgCysProArgProProArgArgSerValThrAlaArgPro 6292
 QY |||:|||||
 740 CGC----- 742
 Db |||:|||||
 6293 ArgSerProSerGlySerArgAlaLeuThrTrpProArgSerGlyThrProTrpPro 6312
 QY |||:|||||
 742 ----- 742
 Db |||:|||||
 6313 IleGlySerProArgTrpSerThrGlyArgProSerTrpArgProAlaGlyThrAsnTrp 6332
 QY |||:|||||
 743 ---CCCGATCGCTCGCGCGCGCGCGCGAGG-----AGCGCGCGAGCGCGG 787
 Db |||:|||||
 6333 ProProGlyCysAlaArgTrpProLeuArgAsnProArgProValSerProGlyPro 6352
 QY |||:|||||
 788 CGGCAAGTCTCCAGCT---CCTTCG-----CCATCG----- 817
 Db |||:|||||
 6353 ArgCysArgMetProAlaValProSerGlyCysSerArgValThrGlyProSerGlyPro 6372

QY 818 -----ACAGCATCTCGGCAAGCCCTTCGCGAGCGCTC----- 850
 Db |||:|||||
 6373 GlyTrpAlaAlaAsnCysTrpLysProSerArgSerProArgSerThrAsnTrpSer 6392
 QY |||:|||||
 851 -----GCTCAGGACACGCGCCCGCGGACGA----- 877
 Db |||:|||||
 6393 ArgCysSerArgArgArgSerGlySerArg---ProGlyArgCysCysTrpArgAlaThr 6411
 QY |||:|||||
 878 ---CGCTTCACTGGGGCGCGCGCTTCGCGCGCTGCGCG----- 916
 Db |||:|||||
 6412 ThrProArgSerThrAlaProArgProSerSerArgCysSerSerGlySerProArgCys 6431
 QY |||:|||||
 917 -----CGTTCCCGC 925
 Db |||:|||||
 6432 GlyGlyHisGlyAlaSerSerProProSerSerAlaThrArgSerValArgSerPro 6451
 QY |||:|||||
 926 CGCTCTCCCGC-----CGCGCGCTGCGAGGCGC-----TGCTGC 961
 Db |||:|||||
 6452 ProProSerProProGlyArgProProThrGlyProGlySerAlaValGlyArgCysCys 6471
 QY |||:|||||
 962 CGCTCTCGCGTACGCGCGCGCGCGCGCGCTGG----- 1000
 Db |||:|||||
 6472 ---CysAlaArgProArgGlyValAlaArgTrpArgTrpSerValCysProSerThrArg 6490
 QY |||:|||||
 1001 -----GCGCGCGAGCGCGAGTGCCACCGACCGCGCGCGCTCTCGCTTG 1048
 Db |||:|||||
 6491 ProProSerGlySerArgAlaMetThrArgTrpSerArgProSerArgPro----- 6507
 QY |||:|||||
 1049 CACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCG----- 1096
 Db |||:|||||
 6508 ---ArgArgArgProProSerProValThrProAlaArgSerArgArgSerSer 6524
 QY |||:|||||
 1097 -----CCGCGC 1102
 Db |||:|||||
 6525 AlaAlaGlyProThrArgGlyTrpSerTyxGlySerProProThrTrpProSerThr 6544
 QY |||:|||||
 1103 GCGCGCACCTGACTGCGCGCGCTGC-----GGCTGCGCGCGCGCGCGCGCTTAG 1156
 Db |||:|||||
 6545 AlaArgThrTrpThr---ProCysSerThrGlyCysAlaArgPro-----ProThr 6560
 QY |||:|||||
 1157 TCCGNCGTCTCGCGCGCGCGCGCGCTGCTGCTACCGGCTGGAGACGCTCC 1201
 Db |||:|||||
 6561 SerValProAlaHisArgThrArgArgSerThrArgArgArgSer 6575

RESULT 14

US-10-084-846A-6
 ; Sequence 6, Application US/10084846A
 ; Publication No. US20040006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MUELENWEG, AGNES
 ; APPLICANT: TREFFZER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; PCT FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: PatentIn Ver. 3.2
 ; SEQ ID NO 6
 ; LENGTH: 19662
 ; TYPE: PRT
 ; ORGANISM: Streptomyces viridochromogenes
 ; FEATURE:
 ; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.
 ; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
 US-10-084-846A-6
 Alignment Scores:

Pred. No.: 3,15e-10 Length: 19662
 Score: 369.00 Matches: 182
 Percent Similarity: 34.64% Conservative: 21
 Best Local Similarity: 31.06% Mismatches: 166
 Query Match: 16.43% Indels: 217
 DB: 15 Gaps: 35

US-10-087-080-31 (1-1209) x US-10-084-846A-6 (1-19662)

QY 1208 CAAGCTAGAGCGTCTCCACCGGTACGACAGTGTCCGGGCA---GGACGCGGACTAAG 1152
 Db 12641 ArgAlaProAlaProAlaGlyGluArgAlaGlyProTTPGlyArgValAlaSer 12660
 QY 1151 GCCGCTCAGCGTTCGGGCGAGCGGAGGCGGAGTACAGTGTCCGGCGCGCGCGCC 1092
 Db 12661 ProAlaArgSerAlaArgAlaProArgAlaProGlyAsnAlaAlaProAlaThr 12680
 QY 1091 GGGCTCGAGTGGCTTGGCGG----- 1070
 Db 12681 AspProArgThr-ThrProArgProArgProAlaArgArgCysProAl 12700
 QY 1069 -----GGGCGCGCGCGGAGAGGTCAAGCAGGAGGCGCGGTGGTGGC 1020
 Db 12700 aArgProAlaAlaArgProGlySerAlaGlyArgAlaGlySerProProGlyValPr 12720
 QY 1019 ACCTCGGCT-----CGCGCGCGCC 999
 Db 12720 oProArgProGlyArgAlaArgCysArgArgAlaArgProAlaArgProArgPr 12740
 QY 998 AGCCGCGCG-----GCTCCCGCGCGGTACCGGAGA-----GC 963
 Db 12740 oSerProProValAlaCysArgArgSerArgArgAlaArgProCysGlyAl 12760
 QY 962 GCGAGCAGCGCTTCAGCGCGCGCGGAGGAGCGCGGGAACG-----CGGCGAGC 909
 Db 12760 aAlaGlyProArgArgPro-----AlaArgSerThrSerArgProAl 12777
 QY 908 GCGGCG-----AGCGCGCGCGCGCTCAAGCTGCTCCCGGCGCC 864
 Db 12777 aAlaGlyProArgThrArgArgAlaArgAlaProGluArgAlaValHisArgSp-- 12796
 QY 863 GTGTCCTGAGCGACGCTCGGAAGGCTTGGCAGAGTGTCTCGATGGCGCAAGGAG 804
 Db 12797 -----ArgArgThrValAlaArgArgSerAsnArgAsnArgSe 12809
 QY 803 CTGAGAACTTCCCGCGGCGCTCGCGCTCTCTGCG-----CGGCGGCGAGCGCATG 747
 Db 12809 rArgValArgProProArgProAlaAlaGlyArgProGlyArgGlyArgProValArgAl 12829
 QY 746 CGGCGAGCGCGGCG 695
 Db 12829 aGlyAlaArgProGlyProAlaAlaThrAlaProProAspArgArgGlyProAlaArgPr 12849
 QY 694 -----GGGCT-----CCTCGGCGCG 678
 Db 12849 oSerGlyProArgHisGlyArgAlaArgArgGlySerGlyArgArgProArgProSe 12869
 QY 677 AGCCGCGCGCGGACCGCGCGC----- 653
 Db 12869 rGlyArgProAlaAlaProProArgProArgGlyArgProGlySerAlaAlaAspProPr 12889
 QY 652 -----GCTGCTCAGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
 Db 12889 oSerAlaProGlyGlySerSerCysProGlyGlyArgGlyAlaLeuArgProArgAr 12909
 QY 611 CGGA----- 608
 Db 12909 gArgProGlyThrProProAlaArgArgProGlyArgSerProArgArgGlyArgG1 12929
 QY 607 -----AGTGTACTCG-----TGTGGGTTGAGCATCTCAGTAGTTG 570
 Db 12929 yCysProGlyThrGlyGlyCysArgArgProAlaCysCysAlaThrAlaAlaGlyArgC1 12949

QY 569 TCCT-----TGCCCGCAGGCGCGG-----AGGGTCGCGCA-- 539
 Db 12949 sProArgArgArgArgAlaCysProAlaAlaCysGlySerArgArgArgGlyArgAlaAr 12969
 QY 538 -----GCACCTTGACGAGCAGTGTGAGCGAAAGGTTGT----- 503
 Db 12969 gProLeuAlaProArgArgTTPAsnAlaThrAlaGlyGlyAlaAlaArgGlyArgAlaAl 12989
 QY 502 -----GGGCGCAGGTTGCGCCAGCCGCTGAGTCCCGCGGAAAGGGGAC 453
 Db 12989 aGlyLeuArgGlySerArgSerAlaGlyAlaProThrSerProArgSerArgGlySe 13009
 QY 452 T-----TGCCCATGAGGTAC 438
 Db 13009 rGlyArgArgProArgArgArgGlyArgSerGlyProArgAlaGlyCysAlaValAlaSe 13029
 QY 437 T-----CGTTGATCTCCGCCAGCGTCAAGCCCGCGCGCGAGTCCGGATGCGCATGGCG 381
 Db 13029 rValArgArgThrProProArgValArgProAlaSerArgGly----- 13045
 QY 380 ATGAGCGCATGTACAGTAGGGGCTTGGCGCGCGGTATATGGCTTGTGCTGTGCTGCA 321
 Db 13046 -----SerSerAlaArgProArgGlyAlaGlyThrAla----- 13056
 QY 320 CCTCTCCGCTCCCG-----CGCGCGCGCGCTTGGCGCGCGCGCGCGCGCTCCGGCGCC 264
 Db 13057 -----ArgArgThrProGlyGlyArgProArg-----ProProArgSerArgArgAr 13072
 QY 263 TCGCCACACCGCTGTCAGCAGCTGCTCGCGGATCGCTCTCT----- 221
 Db 13072 gAlaAlaProGly-----CysAlaGlySerProProSerThrArgArgArgAr 13088
 QY 220 -----CGCGCGCGCG-----CGCTCCCGCACTGTGTCGCGCTGCGCTCGCTATCT 171
 Db 13088 gSerProTTPProAlaGlyProArgAlaPro---SerCysArgAlaAlaCysAlaValAr 13107
 QY 170 CTGG-----CGCGCGCGCGCGCGAGCGCTTG----- 143
 Db 13107 gtrPaspSerAlaArgArgArgProArgAlaAlaGlyArgAlaGlyGlyProCysAlaGl 13127
 QY 142 -----CGCGC-----AGTCCCATCTGAGCCGAG 117
 Db 13127 yCysGlyArgProArgArgArgProAlaValArgProAlaProArgArgSerGl 13147
 QY 116 GAGTGTCTCCCGCGCGCGAGCGCGG-----ACGGCGCTCGCTG 75
 Db 13147 yArgArgProSerArgProArgArgGlyHisArgSerArgTTPThrSerSerArgTh 13167
 QY 74 CGCG-----CGCGCGCTCCAGGTCACTGCTGCTGCTGCTGCGCGCGCGCGAGGAGC 18
 Db 13167 rArgTTPProArgProProArgHisCysProArgAlaAlaArgProArgArgAl 13187
 QY 17 AACACTCCAACTTCA 2
 Db 13187 aAspProSerProSer 13192

RESULT 15
 US-10-029-386-32327
 ; Sequence 32327, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI
 ; FILE REFERENCE: A60MICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 32327
 ; LENGTH: 237

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 29, 2004, 11:57:01 ; Search time 26.5 Seconds
(without alignments)
8777.023 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189

Sequence: 1 atgaagtggagggtgttgcgt.....tggagacgtctctagtgtga 1209

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p -n2p -DEV=xlp
-Q=cpn2.1/USPTO.spool.p/US10087080/runat.29032004.114527.6563/app.query.fasta.1.1351
-DB=PIR_78 -QFMT=fasta -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0
-UNITS=BITS -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080 @CCN 1.1.38 @runat.29032004.114527.6563 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1314.5	60.1	439	2 I60916	HNF-3/forkhead hom
2	543.5	24.8	440	2 S71795	transcription fact
3	517	23.6	465	2 G02738	PREAC-4 - human
4	431.5	19.7	444	2 T09474	forkhead protein F
5	387	17.7	473	2 T03257	forkhead transcrip
6	380.5	17.4	376	2 I49735	hepatocyte nuclear
7	380	17.4	461	2 S34472	MRF-1 protein mo
8	366.5	16.7	468	1 A54258	transcription fact
9	358	16.4	476	2 A54743	transcription fact
10	356	16.3	459	1 B54258	transcription fact
11	356	16.3	466	2 A36674	transcription fact
12	348.5	15.9	354	1 S35090	transcription fact
13	345	15.8	480	2 JH0672	brain factor 1 pro
14	344.5	15.7	458	2 B39533	transcription fact

fork head protein
HBF-C2 (HFK-2) pro
fork head protein
hypothetical prote
transcription fact
FREAC-1 protein - Af
XfKH2 protein - Af
PREAC-2 protein - Af
fork head domain p
nuclear protein fk
BHLFI protein - hu
forkhead protein F
transforming prote
probable transcrip
hypothetical prote
silk gland factor-
slp2 protein - fru
brain factor-2 - r
slippy paired prot
slp2 protein - fru
hypothetical prote
HNF-3/fork-head ho
HNF-3/fork head fa
fork head homolog
fork head transcrip
HNF-3/fork-head ho
HNF-3/fork-head ho
transcription fact
HNF-3/fork-head ho
HNF-3/fork-head ho
forkhead transcrip
HNF-3beta - Africa

ALIGNMENTS

RESULT 1

I60916
HNF-3/forkhead homolog-1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
C:Accession: I60916
R:Clcvidence, D.E.; Overdier, D.G.; Tao, W.; Qian, X.; Pani, L.; Lai, E.; Costa, R.H.
Proc. Natl. Acad. Sci. U.S.A. 90, 3948-3952, 1993
A:Title: Identification of nine tissue-specific transcription factors of the hepatocyte
A:Reference number: A47450; MUID:93248207; PMID:7683413
A:Accession: I60916
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-439 <RES>
A:Cross-references: GB:L13201; NID:g951067; PIDN:AAA74561.1; PID:g550513
C:Genetics:
A:Gene: HNF-1
A:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
P100-192/Domain: fork head DNA-binding domain homology <PHD>

Alignment Scores:
Pred. No.: 4.42e-58 Length: 439
Score: 1314.50 Matches: 278
Percent Similarity: 74.30% Conservative: 14
Best Local Similarity: 70.74% Mismatches: 86
Query Match: 60.05% Indels: 15
DB: 2 Gaps: 6

US-10-087-080-31 (1-1209) x I60916 (1-439)

QY 49 GGCGAGTACCTGGAGGGCGGGCGGCGAGCGCGCTCCCGCTGTGCGCGCGGGA 108
DB 2 GlySerAspLeuGluGlyAlaGlySerSerAspValProSerProLeuSerAlaAlaGly 21
QY 109 GACGACCTCCCTGGGCTCAGATGGGAGCTGCGCGGCC---AAGCGTCCGCGGGCGGGCC 165
DB 22 AspAspSerLeuGlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlySerGly 41


```
Db      270  lnpProProCysAlaTyTyGlyProTyTyGlyCysGlyTyTyGlyLeuGlnLeuGlnProTyTyH 290
Qy      765  CCAGGAGGAGCGG-----CCAGCCCGGGGCAAGTTCTCCAGCTC 806
Db      290  isProHisSerAlaLeuPheAlaPheHisPheSerProProProProProProProProProProPro 310
Qy      807  CTTGCGCATCGACAGCATCTCGCGAAGCCCTTCGCGACCGCTCGCTCAGGACACAGCG 866
Db      310  laAlaProAlaGlyAla-ProAlaAlaAlaLeuProProProProProProProProProProPro 329
Qy      867  CCGCGGAGCAGCGCTTCACTGGGGCGCCCGCC-----CTCGCGCGCT-----GCC 914
Db      330  ProArg-----ArgArgAlaProLeuLeuProAlaAlaGlnLeuAla 343
Qy      915  CGGTTCCCGCGCTCTCCCGCGCGCCCTGCAGGGCCCTGCGCGCTCTGCGCGTA 974
Db      344  ArgThProPheGlyTyTyProHisProLeu-----GlyProAlaAlaAlaSerLeu 361
Qy      975  CGCGCGCGGAGCGCGCGCGCTGGCGCGCGCGAGGCGGAGTGCACCGCGCC 1034
Db      362  HisAlaAlaLysProGlySer-----GlyAlaAlaValAlaArg 374
Qy      1035  GCCCTCTCTGCTGCACTCT-----CCGCGGGGGCGCCCGCAA-----GCCACT 1082
Db      375  SerProPheSerIleGluSerIleIleGlyGlyProGlyProGlyLeuGlyAlaGly 394
Qy      1083  CCGAGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
Db      395  ProAlaProGlyAlaGlyGlySerCysAlaSerGlnSerGlyAlaAlaThrGlyLeuSer 414
Qy      1134  CGAGCGCTGCGAGCGCGCTTGTAGTCGCGCTGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1190
Db      415  ArgSerLeuGlySerGlyLeuAlaProAlaAla---AlaLeuProAlaAlaProGly 432

RESULT 3
G02738
FREC-4 - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 19-May-2000
C:Accession: G02738
R:Enetback, S.
submitted to the EMBL Data Library, June 1996
A:Reference number: H01646
A:Accession: G02738
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-465 <ENE>
A:Cross-references: EMBL:U59832; NID:g1399238; PID:g1399239
C:Genetic:
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:125-216/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.:      1,12e-18      Length:      465
Score:          517.00      Matches:     157
Percent Similarity: 44.21%      Conservative: 34
Best Local Similarity: 36.34%      Mismatches: 132
Query Match:      23.62%      Indels:     109
DB:              2          Gaps:        16

US-10-087-080-31 (1-1209) x G02738 (1-465)

Qy      60  GGAGGCGCGGCGGCGGAGCGCGCGCTGCTGCGCGCGGAGAGCGACTCCCT 119
Db      39  GlyGlyGlyGlyProArgLeuAlaValPro----- 49
Qy      120  GGGCTCAGATGGGAGCTGCGCGCCCAAGCGCTCCGCGCGCGCGCGCGAGATACGA 179
Db      50  -----AlaGlnArgArgArgSerTyAla 60
Qy      180  GGGCAGCGCGGACAGAG-----TGGGG 203
```

```
Db      61  GlyGluAspGluLeuGluAspLeuGluGluGluAspAspAspAspAspAspAspAspAspAspAsp 80
Qy      204  AGGCGGCGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 263
Db      81  ProProAlaGlyGlySerProAlaProProGlyProAlaProAlaProAlaProAlaProAlaProAla 100
Qy      264  GGGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 323
Db      101  GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 119
Qy      324  ACGCAGCAGCATATACGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCG 383
Db      119  aLysAsn---ProLeuVal-----LysProProTyTySerTyTyIleAlaLeuIleTh 135
Qy      384  CATGGCCATCGGACCTCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 443
Db      135  rMetAlaIleLeuGlnSerProLysLysArgLeuThrLeuSerGluLeuCysGluPheI 155
Qy      444  CATGGGCAAGTTCCCTTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 503
Db      155  eserGlyArgPheProTyTyArgGluLysPheProAlaTrpGlnAsnSerIleArgH 175
Qy      504  CAACCTTTCTCAACGACTGCTTCAAGGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 563
Db      175  sAsnLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro---G 194
Qy      564  CAAGGACAACCTACTGATGCTCAACCCACAGCAGTACACCTTCCGCCGCGGCGGCTT 623
Db      194  YLysGlyAsnTyTrpThrLeuAspProGluSerAlaAspMetPheAspAspGlySerPh 214
Qy      624  CCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 683
Db      214  eLeuArgArgLysArgPheLys---ArgGlnProLeuLeuProProAsnAlaAla 233
Qy      684  CGAGGAGGCC-----CCGCGCGCTCCCGCGCGC 710
Db      233  aAlaGluSerLeuLeuArgGlyAlaGlyAlaAlaGlyGlyAlaGlyAspProAlaAla 253
Qy      711  CCGG-----CCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 742
Db      253  aAlaAlaAlaLeuPheProProAla-ProProProProHisAlaTyTyGlyTy 273
Qy      743  -----CCGCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
Db      273  roTyTyGlyCysGlyTyTyGlyLeuGlnLeuProProTyTyAlaProProSerAlaLeuPheA 293
Qy      774  GCGCGCGCGCGCGCGGCGGCAAGTTCTCCAGCTCTCCGCGCATCGACAGCATCTCGCGCGCA 833
Db      293  laAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaPheHisProHis-----S 308
Qy      834  GCCCTTCGCGCGCGCTCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
Db      308  erPro-ProProProProPro---HisGly----- 317
Qy      894  CGCGCGCTCCCGCGCGCTGCGCGCGGTTCCCGCGCT-----CCT 932
Db      318  -----AlaAlaAlaGluLeuAlaAlaArgThrAlaPheGlyTyTyArgProHisPro 333
Qy      933  CCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 992
Db      334  LeuGlyAlaAlaLeuProGlyProLeuProAlaSerAlaAlaLysAlaGlyGlyProGly 353
Qy      993  CGCGCTGGGCGCGCGCGAGGC-----CGAGTGCACCGC 1028
Db      354  AlaSerAlaLeuAlaArgSerProPheSerIleGluSerIleGlyGlySerLeuGly 373
Qy      1029  CGCGCGCGCGCTCTGCTGCGCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1088
Db      374  ProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 393
Qy      1089  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1148
Db      394  Pro-----SerProValAlaAlaProProAlaProGlySerSerGlyGly 408
```

```
QY 1149 GGCCTTAGTCGCGCTGCTGCGCCGACCT 1178
    |||  ::  |||
Db 409 GlyCysAlaAlaGlnAlaAlaValGlyPro 418
    |||  ::  |||

RESULT 4
T09474
forkhead protein FRAC-2 - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-May-2000
C:Accession: T09474
R:Hellqvist, M.; Mahlapuu, M.; Blixt, C.; Enerback, S.; Carlsson, P.
J. Biol. Chem. 273, 23335-23343, 1998
A:Title: The human forkhead protein FRAC-2 contains two functionally redundant activator
A:Reference number: Z16882; MUID:9839768; PMID:9722567
A:Accession: T09474
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-444 <HEL>
A:Cross-references: EMBL:U13220; NID:G3425849; PID:G3425850
A:Experimental source: lung
C:Function:
A:Description: may function as transcription activator
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
F:100-191/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.: 1-9e-14 Length: 444
Score: 431.50 Matches: 129
Percent Similarity: 42.62% Conservative: 41
Best Local Similarity: 32.49% Mismatches: 156
Query Match: 19.71% Indels: 69
DB: 2 Gaps: 11

US-10-087-080-31 (1-1209) x T09474 (1-444)

QY 73 GCGAGCGACCGCGCTCCCGCTGTCGGCGGGGAGACGACTCCCTGGCTCA---GAT 129
    |||  ::  |||
Db 5 GlyGlyProProAlaProLeuArgAlaCysSerProValProGlyAlaLeuGln 24
    |||  ::  |||

QY 130 GGGGACTCGCGCCCAAGCGCTCGCGGGCGGGCGCCAGAGATACGACGGCGCGGC 189
    |||  ::  |||
Db 25 AlaAlaLeuMetSerProProAlaAlaAlaAlaAlaAlaAlaProGluThr 44
    |||  ::  |||

QY 190 GAACAGATCGGGAGCGGGCGGGCGGGAGAGCGATCCGCGACGACTGCTGCA 249
    |||  ::  |||
Db 45 ThrSerSerSerSerSerSerAlaSerCysAlaSerSerSerSerSerSer 64
    |||  ::  |||

QY 250 GCGGTGTCGCGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGG 309
    |||  ::  |||
Db 65 SerAlaSerAlaProSerAlaAlaCysLysSerAlaGlyGlyAlaGlyAlaGly 84
    |||  ::  |||

QY 310 AGCGGCGAGGTGCGACGACGACGACGACGACGACGACGACGACGACGACGAC 366
    |||  ::  |||
Db 85 SerGlyAlaLysLysAlaSerSerGlyLeuArgArgProGluLysProProTyrSer 104
    |||  ::  |||

QY 367 TACATCGCTCATCGCTATCGCTATCGCTATCGCTATCGCTATCGCTATCGCTAT 426
    |||  ::  |||
Db 105 TyrAlaAlaLeuValMetAlaLeuGlnSerSerProSerLysArgLeuThrLeuSer 124
    |||  ::  |||

QY 427 GAGATCAACGAGTACCTCATGCGCAAGTTCCTCTTTCGCGGCGAGCTACACGGCTGG 486
    |||  ::  |||
Db 125 GluIleTyrGlnPheLeuGlnAlaArgPheProPheArgGlyAlaTyrGlnGlyTrp 144
    |||  ::  |||

QY 487 CGAATCGCTGCGCACACCTTCGCTACGACTGCTTCGCTACGACTGCTGCGCGAC 546
    |||  ::  |||
Db 145 LysAsnSerValArgHisLeuSerLeuAsnGluCysPheIleLysLeuProLysGly 164
    |||  ::  |||

QY 547 CCCTCGCGCGCTCGGGGACGAGCACTACTGATGCTCAACCCCAACGAGGATACAC 606
    |||  ::  |||
Db 165 LeuGlyArgPro---GlyLysGlyHisTyrTrpThrIleAspProAlaSerGluPheMet 183
    |||  ::  |||

QY 607 TTGCGCGACGGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
    |||  ::  |||
```

```
Db 184 PheGluGlySerPheArgArgArgProArgGlyPheArgArgLysCysGlnAlaLeu 203
    |||  ::  |||
QY 667 GCGCCC-----GGGCTG-----
    |||  ::  |||
Db 204 LysProMetTyrHisArgValValSerGlyLeuGlyPheGlyAlaSerLeuProGln 223
    |||  ::  |||
QY 679 -----CGCGCCGAGGAGGCGCCG-----
    |||  ::  |||
Db 224 GlyPheAspPheGlnAlaProProSerAlaProLeuGlyCysHisSerGlnGlyTyr 243
    |||  ::  |||
QY 697 ---GGCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
    |||  ::  |||
Db 244 GlyGlyLeuAspMetMetProAlaGlyTyrAspAlaGlyAlaGlyAlaProSerHisAl 263
    |||  ::  |||
QY 754 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
    |||  ::  |||
Db 263 AHISProHisHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 283
    |||  ::  |||
QY 814 ATCGACGACGATCTCTGCGCAAGCCCTTCGCGACGCGCTCGCTCAGGGACACGCGCGCGG 873
    |||  ::  |||
Db 283 rTyrMetAlaSerCysProValProAlaGly-----ProGlyG 296
    |||  ::  |||
QY 874 ACGAGCTTCTAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
    |||  ::  |||
Db 296 yValGlyAlaAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 314
    |||  ::  |||
QY 934 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 993
    |||  ::  |||
Db 314 rSerProValProSerSerSerProAlaMetAlaSerAla----- 326
    |||  ::  |||
QY 994 CGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1049
    |||  ::  |||
Db 327 -----IleGluCysHisSerProTyrThrSerProAlaAlaHisT 340
    |||  ::  |||
QY 1050 --ACCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1107
    |||  ::  |||
Db 340 rpSerSerProGlyAlaSerProTyrLeuLysGlnPro----- 352
    |||  ::  |||
QY 1108 CACTGTACTGCCCCCTGCGGCTGCCCGCGACGCGCTGCGAGCGCGCTT 1154
    |||  ::  |||
Db 353 --ProAlaLeuThrProSerSerAsnProAlaAlaSerAlaGlyLeu 367
    |||  ::  |||

RESULT 5
S70357
forkhead transcription factor HNF-3 alpha - human
N:Alternate names: hepatocyte nuclear factor-3 alpha
C:Species: Homo sapiens (man)
C>Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 16-Jul-1999
C:Accession: S70357
R:Bingle, C.D.; Gowan, S.
Biochim. Biophys. Acta 1307, 17-20, 1996
A:Title: Molecular cloning of the forkhead transcription factor HNF-3-alpha from a human
A:Reference number: S70357; MUID:96254057; PMID:8652662
A:Accession: S70357
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-473 <BIN>
A:Cross-references: EMBL:U39840; NID:gl066121; PID:AB06493.1; PID:gl066122
A:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
F:171-262/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.: 2.97e-12 Length: 473
Score: 387.00 Matches: 124
Percent Similarity: 44.36% Conservative: 45
Best Local Similarity: 32.55% Mismatches: 132
Query Match: 17.68% Indels: 80
DB: 2 Gaps: 15

US-10-087-080-31 (1-1209) x S70357 (1-473)

QY 70 GCGCGCGACGCGCGCGCTCCCGCTGTCGGCGGGGAGACGACTCCCTGGGCTCAGAT 129
    |||  ::  |||
```

```

88 GlySerAlaGlyAlaMetAsnSerMetThrAlaAlaGlyValThrAlaMetGlyThr--- 106
130 GGGGACTGCGCGCAAGCGTCCCGCGCGCGCGCCAGATACGACGCGGCGAGCGC 189
107 -----AlaLeuSerProSer---GlyMetGlyAlaMetGlyAlaGlnAlaAla 122
190 GAACAGAGTGGGAGGCGGCGCGCGCGGAGGAGGATCCCGGACAGCTGCTGCA 249
123 SerMetMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCysMetSerProMet 142
250 GCGGTGGTGGCGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 309
143 AlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGly-----GlyGlyGlyAsp 159
310 ACCGCGGAGGTGACGACGAGCAGCATATACCGCGCGCGCGCGCGCGCGCGCGCG 369
160 AlaTyrThrPheLeuArgSerTyrProHisAla-----LysProProTyrSerTyr 176
370 ATCGCGCTCATCGCATCGCATCGCATCGCATCGCGCGCGCGCGCGCGCGCGCGCG 429
177 IleSerLeuIleThrMetAlaIleGlnArgAlaProSerLysMetLeuThrLeuSerGlu 196
430 ATCAACAGTACTCATGCGCAAGTTCCTTTCCTTCCTTCCTTCCTTCCTTCCTTC 489
197 IleTyrGlnTyrIleMetAspLeuPheProTyrTyrArgGlnAsnGlnArgTyrGln 216
490 AACTCGGTGGCGCAACCTTCGTCAACGACTGCTTCGTCAAGGTGCTCGCGGACCC 549
217 AsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSerPro 236
550 TCGCGGCTCGCGGCAAGCAACTACTGATGCTCAACCCCAACAGCAGTACACCTTC 609
237 AspLysPro---GlyLysGlySerTyrTyrTrpThrLeuHisProAspSerGlyAsnMetPhe 255
610 GCCGACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
256 GluAsnGlyCysTyrLeuArgGlnLysArg----- 266
670 CCGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
267 -----PheLysCysGluLysGlnProGly-----AlaGlyGly 277
730 GCCCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783
278 GlyGlyLysGlySerGlyLysGlySerGlyAlaLysGlyProGluSerArgLysAsp 297
784 CCGCGGCAAGTTCCTCGAGTCTCTCGCATCGACAGCATCTCGCGCAAGCCCTTCGCG 843
298 ProSerGly-----AlaSerAsnProSerAlaAspSerProLeuHisArgGlyValHis 315
844 AGCGTGCCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 903
316 -----GlyLysThrGlyGlnLeuGluGlyAlaProAla 326
904 CCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 963
327 ProGlyProAlaAlaSerProGlnThrLeuAsp----- 337
964 CTCTGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1023
338 -----HisSerGlyAlaThrAlaThrGlyGlyAlaSerGluLeuLysThrPro 353
1024 -----CCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062
354 AlaSerSerThrAlaProProIleSerSerGlyProGlyAlaLeuAlaSerValProAla 373
1063 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122
374 SerHisProAlaHisGlyLeu---AlaProHisGluSerGlnLeuHisLeu----- 389
1123 CTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1182

```

```

390 -----LysGlyAspProHisTyrSer 396
1183 TAC 1185
397 Phe 397
RESULT 6
hepatocyte nuclear factor 3 forkhead homolog 8 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 24-Sep-1999
C/Accession: I49735
R/Clevidence, D.E.; Overdier, D.G.; Peterson, R.S.; Porcella, A.; Ye, H.; Paulson, K.E.;
Dev. Biol. 166, 195-209, 1994
A/Title: Members of the HNF-3/forkhead family of transcription factors exhibit distinct
A/Reference number: I49735; MUID:95046902; PMID:7958446
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-376 <RES>
A/Cross-references: GB:I35949; NID:9762833; PIDN:AAA64885.1; PID:9575512
C/Genetics:
A/Gene: HNF-8
C/Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F;47-138/Domain: fork head DNA-binding domain homology <FHD>
Alignment Scores:
Pred. No.: 6,48e-12 Length: 376
Score: 380.50 Matches: 112
Percent Similarity: 44.65% Conservative: 34
Best Local Similarity: 34.25% Mismatches: 100
Query Match: 17.38% Indels: 81
DB: 2 Gaps: 12
US-10-087-080-31 (1-1209) x I49735 (1-376)
219 GGAGGAGGAGTATCCCGGACAGCTGCTGACAGGGTGTGGCGGAGGCGCGGAGCGCGG 278
12 GlyGlyGly-----ThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 24
279 GCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
25 -AlaMetAspProAlaAlaAlaProHisGlnGlnGluAspGlnArgArg----- 42
339 TAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398
43 -AlaArgProGluLysProProTyrSerTyrIleAlaLeuIleValMetAlaIleGlnSe 62
399 CTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458
62 rSerProSerLysArgLeuThrLeuSerGluIleTyrGlnPheLeuGlnAlaArgPhe 82
459 CTCTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 518
82 oPhePheArgGlyAlaTyrGlnGlyTyrLysAsnSerValArgHisAsnLeuSerLeuAs 102
519 CGACTGCTTCGTCAAGTGTCTGCGGACCCCTCGCGCGCGCGCGCGCGCGCGCGCG 578
102 nGluCysPheIleLysLeuProLysGlyLeuGlyArgPro---GlyLysGlyHisTyrTr 121
579 GATGCTCAACCCCAACAGCGAGTACACCTTCGCCGACGGGTCTTCGCCGCGCGCGCA 638
121 pThrIleAspProAlaSerGluPheMetPheGluGluGlySerPheArgArgProAr 141
639 GCGCTCAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
141 gGlyPheArgArgLysCysGlnAlaLeuLysProValTyrSerMetValAsnGlyLeuGl 161
681 GCCCGAGGAGGCGCGCG-----GGCTCCCGCGCGCGCGCGCGCGCGCGCGCG 713
161 yPheAsnHisLeuProAspThrTyrGlyPheGlnGlySerGlyGlyLeuSerCysAlaPr 181
714 GCCG----- 717

```


A:Residues: 1-468 <KAE>
A:Cross-references: EMBL:X74936; NID:G402194; PIDN:CAA52890.1; PID:G402195
F:Saeki, H.; Hogan, B.L.
Development 118, 47-59, 1993
A:Title: Differential expression of multiple fork head related genes during gastrulation
A:Reference number: I49674; MUID:93387221; PMID:8375339
A:Accession: I49675
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 'XX',161-269 <RES>
A:Cross-references: GB:L10408; NID:G404761; PIDN:AAA03160.1; PID:G404762
C:Genetics:
A: Introns: 24/3
C:Function:
A:Description: regulation of gene expression, especially in liver and lung
A:Note: expressed in embryonic endoderm and endoderm-derived adult tissues, most strongly
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
C:Keywords: DNA binding; transcription factor
F:170-261/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:	
Pred. No.:	3,078-11
Score:	386.50
Percent Similarity:	41.25%
Best Local Similarity:	30.23%
Query Match:	16.74%
DB:	1
	Length: 468
	Matches: 116
	Conservative: 42
	Mismatches: 128
	Indels: 97
	Gaps: 14
US-10-087-080-31 (1-1209) x A54358 (1-468)	

124	TGATGGGAGTCTGGCGGCCAAGCCGTCCGGCGGGCGCGCCAGAGATACGCAGGCG	183
Qy		
78	SerProGlyAlaValAlaGlyMetProGlyAlaSerAlaGlyAlaMetAsnSerMetThr	97
Db		
184	GACGGCAACAGAGTGCGGAGGC-----GGCGCGGCGCGGAGGAGCGATCCCG	234
Qy		
98	AlaAlaGlyValThrAlaMetGlyThrAlaLeuSerProGlyGlyMetGlySerMetGly	117
Db		
235	GNAGCAGCTGTCAGAGCGGTGGCGGAGCGGCGGAGCGCGGCGCGCGGGCCA---	291
Qy		
118	AlaGlnProValThrSerMetAsnGlyLeuGlyProTyrAlaAlaAlaMetAsnProCys	137
Db		
292	-----GGCGCGGCGCGCGGCGCGGCGGCGGCGGCGAGCGCGAG	318
Qy		
138	MetSerProMetAlaTyrAlaProSerAsnLeuGlyArgSerArgAlaGlyGlyGly	157
Db		
319	GGTGA-----CGCAGCAGCCATATACGGCGGCGGCCCAAGCCCGCTACTCG	366
Qy		
158	AspAlaLysThrPheLysArgSerTyrProHisala-----LysProProTyrSer	174
Db		
367	TGATCGCGCTCATCGCATGCCATCGCACTCGCGACTCGCGCGGCGCGCTTGACGCTGGC	426
Qy		
175	TyrIleSerLeuIleThrMetAlaIleGlnAlaProSerLysMetLeuThrLeuSer	194
Db		
427	GAGATCAACGAGTACTCATGGCAAGTCCCTTTTCGCGCGCAGCTACACGGCTCG	486
Qy		
195	GlutIleTyrGlnTrpIleMetAspLeuPheProTyrArgGlnAsnGlnArgTrp	214
Db		
487	CGCACTCGCTGGCGGCACAACTTTGGCTCAACGACTGCTTGTCGAAGGTGCTGGCGAC	546
Qy		
215	GlnAsnSerIleArgHisSerLeuSerPheAsnAspCysPheValLysValAlaArgSer	234
Db		
547	CCCTCGCGCCCTGGGGCAAGCAACTACTCTGGATGCTCAACCCCAACGACGAGTACAC	606
Qy		
235	ProAspLysPro--GlyLysGlySerTyrTrpThrLeuHisProAspSerGlyAsnMet	253
Db		
607	TTTCGCGCAGGGGTCTTCGGCGGCGCGCAGCGGCTCAGCCACGCGCGCGCGTCCCG	666
Qy		
254	PheGluAsnGlyCysTyrLeuArgGlnLysArg-----	265
Db		
667	GGCGCCGGGCTCGCGCCCGAGAGGCCCGGGCCCTCCCGCCGCCGCCCGCGCGCC	726
Qy		
266	-----PheLysCysGluLysGlnProGly-----	273
Db		

Qy	727	GC	CCGCCCGCGCTGCGCCCGCCGATCGCGCTCCCGCGCCCGCCGAGAGAGCGCGCCAGCGCCC	786
Db	274	Ala	adlyglyglySerGlyglyGlySerLysGlyglyProGluHisArgLysAspPro	293
Qy	787	GC	GGCAAGTTCTCCAGCTCCTTCGCGATCGACAGCATCTGGCAAGCCCTCCGCGAGC	846
Db	294	Ser	gly-----	295
Qy	847	CT	CGCTCCTCAGGACACGGCCCGCGG-----ACGACGCTTCAGTGGGC	891
Db	296	-----	ProGlyAsnProSerAlaGluSerProLeuHisArgGly	308
Qy	892	GCC	---GCGCCCTGCGCGCTGCCCGCTTCCTCCCGCGCGCGCCCTGC	948
Db	309	Val	HisGlyLysAlaSerGlnLeuGluGlyAlaProAlaProGlyProAlaAlaSer---	322
Qy	949	AGG	CCCTGTCGCGCTCTCGCGGTACGGCGCGCGCGGAGCGCGCGCTGGCGCGCGC	1008
Db	328	-----	ProGlnThrLeuAspHisSerGlyAlaThrAlaThrGlyGlyAlaSer	343
Qy	1009	GAG	CGCGAGGTGCCA-----CGACCGCGCGCCCTCTCTGTTCACCT-----CTC	1056
Db	344	Glu	LeuLysSerProAlaSerSerSerAlaProProIleSerSerGlyProGlyAlaLeu	363
Qy	1057	CCG	CGGCGCGCCCGCCCAAGCACTCCGAGGCGCGCGCGCGCGCGCGCGCGCTAC	1116
Db	364	Ala	SerValProProSerHisProAlaHisGlyLeuAlaProHisGluSerGlnLeuHis	383
Qy	1117	TG	CCCCCTCGCGTGCCTCCGACGCGCTTCGACGGCGCTTAGTCGNCGTCTCTGCCCGCGAC	1176
Db	384	Leu	-----LysGlyAspProHis	389
Qy	1177	CT	GTGTGTAC	1185
Db	390	Tyr	SerPhe	392
RESULT 9				
A54743				
transcription factor HFK1 - human				
C/Species: Homo sapiens (man)				
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 12-Sep-1997				
C/Accession: A54743				
R/Murphy, D.B.; Wiese, S.; Burfeind, P.; Schmoldt, D.; Matti, M.G.; Schulz-Sch				
Genomics 21, 551-557, 1994				
A/Title: Human brain factor 1, a new member of the fork head gene family.				
A/Reference number: A54743; PMID:95048332; PMID:7959731				
A/Accession: A54743				
A/Status: preliminary; nucleic acid sequence not shown				
A/Molecule type: mRNA				
A/Residues: 1-476 <MUR>				
A/Cross-references: GB:X74142				
C/Genetics:				
A/Gene: GDB:FKHL4; HBF-1; HFK1				
A/Cross-references: GDB:433550				
A/Map position: 14q12-14q22				
C/Superfamily: unassigned fork head proteins; fork head DNA-binding domain homo				
F:169-260/Domain: fork head DNA-binding domain homology <FHD>				

Alignment Scores:		
Pred. No.:	8,058-11	Length:
Score:	358.00	Matches:
Percent Similarity:	3.3%	Conservative:
Best Local Similarity:	29.2%	Mismatches:
Query Match:	16.3%	Indels:
DB:	2	Gaps:
		476
		122
		151
		102
		17

US-10-087-080-31 (1-1209) x A54743 (1-476)

Qy		22	CCTCGCGGGCCACGCGGCAGTGA	CCTGGAGGGCGGGCGGACGCGAC	81
Db		68	ProArgAla[ag]n[ng]n-----	-----Pro	76


```
QY 82 GCGCGCTCCCGCTGTGCGCGCGGAGACGACTCCCTGGCTCAGATGGGAGTGGCG 141
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 ProProProProLeuAla----- 82
QY 142 GCCAAGCGCTCGCGCGCGCGGCGCCAGAGATACGACGCGCGACGCGACAGAGTGGC 201
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 -----ProGlnAlaGlyGlyAlaAlaGlnSerAsnAspGluLysGlyProGlnLeuLeu 100
QY 202 GGAGCGCGCGCGCGCGGAGGCGCATCCCGCAGCAGCTGCTGCAGCGGTGGCG 261
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 LeuLeuProProThrAspHisHisArgProProSerGlyAlaLysAlaGlyGlyCys 120
QY 262 GAGCGCGCGAGCGCGCGCGCGCGCGCA-----GCGCGCGCGC 300
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 ArgProGlyGluLeuGlyProValGlyProAspGluLysGlyAlaGlyAlaGly 140
QY 301 GGC-----GCGGAGCGCGCGGAGGTGCACG----- 327
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 GlyGluGluLysGlyAlaGlyGluGlyGlyLysAspGlyGlyLysGluGly 160
QY 328 AGCAAGCATATACGCGCGCGCGCGCAAGCCCTACTCGTACATCGGCTCATCGCATG 387
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 GluLysLysAsnGlyLysTyrgluLysProPheSerTyrgluLysAsnAlaLeuMetMet 180
QY 388 GCCATCCGCGACTCGCGCGCGCGGCGCTTGACGCTGCGGAGATCAACGAGTACCTCATG 447
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181 AlaIleArgGlnSerProGluLysArgLeuThrLeuAsnGlyIleTyrgluPheIleMet 200
QY 448 GGCAAGTTCCTCTTTCCCGCGCGAGCTACACGGCTGGCGCAACTCCGTCGCGCAAC 507
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 LysAsnPheProTyrgluLysGluLysGlnGlyTrpGlnAsnSerIleArgHisAsn 220
QY 508 CTTTCGCTCAACGACTCTCTGTCAGGTGCTGCGGACCCCTCGCGCGCTCGGCGCAAG 567
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 LeuSerLeuAsnLysCysPheValLysValProArgHisTyrgluAspPro---GlyLys 239
QY 568 GACAACCTACTGGATGCTCAACCCCAACAGAGAGTACACTTCGCCGCGGCTTCGCG 627
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 GlyAsnTyrgluMetLeuAspProSerSerAspValPheIleGlyThrThrGly 259
QY 628 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 LysLeuArgArgArgSerThrThrSerProAlaLysLeuAlaPheLysArgGlyAlaAla 279
QY 661 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 LeuThrSerThrGlyLeuThrPheMetAspAlaProAlaProSerThrGlyProCysArg 299
QY 703 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 ProSerCysProCysThrThrProAlaSerSerThrLeuSerTyrgluThrThrSer 319
QY 733 -----CGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 786
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
320 AlaTyrgluProSerHisProMetProTyrgluSerSerValLeuThrGlnAsnSerLeuGlyAsn 339
QY 787 GCGGCAAGTTCCTCACTCC-----TTGCGCATGACAGCATCTCGCGCAAG----- 834
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
340 AsnHisSerPheSerThrAlaAsnGlyLeuSerValAspArgLeuValAsnGlyGluIle 359
QY 835 CCTTCGCGAGCGCTGCTCAGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 ProTyrgluAlaHisLeuThrAlaAlaAlaLeuAlaAlaSerValProCysGlyLeu 379
QY 892 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 951
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 SerValProCys-----SerGlyThrTyrgluSerLeuAsnProCysSer 393
QY 952 GCCTCTGCTG-----CCGCTCTGCGCGTAC 975
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 ValAsnLeuLeuAlaGlyGlnThrSerTyrgluPheProHisValProHisProSerMet 413
QY 976 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1035
```

```
Db 414 ThrSerGlnSerSerThrSerMetSerAlaArgAlaAlaSer---SerSerThrSerPro 432
QY 1036 CCCTCTCTCTGTCACCTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1086
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 Pro-----AlaProArgPro-----LeuProCysGluSerLeuArg 444
RESULT 10
B54258
transcription factor HNF-3 beta - mouse
NAlternate names: hepatocyte nuclear factor 3 beta
CSpecies: Mus musculus (house mouse)
CDate: 18-Aug-1995 #sequence revision 19-Oct-1995 #text_change 18-Jun-1999
CAccession: B54258; I49676; S37185
R/Kaestner, K. H.; Hiemisch, H.; Luckow, B.; Schutz, G.
Genomics 20, 377-385, 1994
A>Title: The HNF-3 gene family of transcription factors in mice: gene structure, cDNA se
AReference number: A54258; MUID:94307723; PMID:8034310
AAccession: B54258
A.Molecule type: mRNA
A.Residues: 1-459 <KAB>
A.Cross-references: EMBL:X74937; NID:9402190; PIDN:CAA52891.1; PID:9402191
R/Sasaki, H.; Hogan, B.L.
Development 118, 47-59, 1993
A>Title: Differential expression of multiple fork head related genes during gastrulation
AReference number: I49674; MUID:93387221; PMID:8375339
AAccession: I49676
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-9, 'H', 11-459 <RES>
A.Cross-references: GB:L10409; NID:9404763; PIDN:AAA03161.1; PID:9404764
C:Genetics:
A.Gene: HNF-3beta
A.Introns: 23/3
C:Function:
A.Description: regulation of gene expression, especially in liver and lung
A.Note: expressed in embryonic endoderm and endoderm-derived adult tissues, most strongly
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
C:Keywords: DNA binding; transcription factor
F:159-250/Domain: fork head DNA-binding domain homology <FHD>
Alignment Scores:
Pred. No.: 1,02e-10 Length: 459
Score: 356.00 Matches: 120
Percent Similarity: 41.49% Conservative: 36
Best Local Similarity: 31.91% Mismatches: 148
Query Match: 16.26% Indels: 72
DB: 1 Gaps: 11
US-10-087-080-31 (1-1209) x B54258 (1-459)
QY 49 GCGAGTGCCTGGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGA 108
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 GlySerGlyAsnMetSerAlaGlySerMetAsnMetSerTyrgluAlaGlyMet 73
QY 109 GACGACTCCCTGGGCTCAGATGGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 168
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 SerProSerLeu-----AlaGlyMetSerProGlyAlaGlyAlaMetAla 88
QY 159 AGAGATACGAGGCGAGCGCGGAAACAGAGT---CGCGGAGCGCGCGCGCGCGCGGAG 225
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 GlyMetSerGlySerAlaGlyAlaAlaGlyValAlaGlyMetGlyProHisLeuSerPro 108
QY 226 CGCATC-----CCGCGCAGCAGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 SerLeuSerProLeuGlyGlyGlnAlaAlaGlyAlaMetGlyGlyLeuAlaProTyrglu 128
QY 271 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129 AsnMetAsnSerMetSerProMetTyrgluGlnAlaGlyLeuSerArgAlaArgAspPro 148
QY 331 AAGCATATACGCGG-----CGGCCAAGCCCCCTACTCGTACATCGCGCTCATC 381
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

Db      374 AlaProHisGluSerGlnLeuHisLeuLysGlyAlaProHisTySerPheAsnHisPro 393
Qy      808 TTCGCATCGACAGATCTTCGCGAAGCCCTTCGCGACCCGCTCGCTCAGGACACGCGCC 867
Db      394 PheSerIleAsnAsnLeuMetSerSerGluGlnGlnHisLeuAspPheLysAla 413
Qy      868 CCGCGACACGCTTCAGTGGGGCGCGCGCCCTCGCCGCGCTCGCCGCTTCGCCGCG 927
Db      414 TyrGluGlnAlaLeuGlnTyrSer 421
Qy      928 CTCCTCCCGCGCGCCCTGCGAGGCCCTGCTGCGCTCGCGCTAGCGCGCGCGCGAG 987
Db      422 -----ProfyrrGlyAlaThrLeu 427
Qy      988 CCGCGCGCGCTGCGCGCGCGAGCGCGAGGTGCCACCGACGCGCGCG-----CCCCTC 1041
Db      428 ProAlaSerLeuProLeuGlyGlyAlaSerValAlaThrArgSerProIleGluProSer 447
Qy      1042 CTGCTTGACCT 1053
Db      448 AlaLeuGluPro 451

RESULT 12
transcription factor HNF-3 gamma - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S35090; A39533
R:Lai, E.; Prezioso, V.R.; Tao, W.; Chen, W.S.; Darnell, J.E.
submitted to the EMBL Data Library, January 1993
A:Description: Hepatocyte nuclear factor 3a belongs to a gene family in mammals that is
A:Reference number: S35090
A:Accession: S35090
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-354 <LA1>
A:Cross-references: EMBL:L09648; NID:G204624; PIDN:AAA41339.1; PID:G204625
R:Lai, E.; Prezioso, V.R.; Tao, W.; Chen, W.S.; Darnell Jr., J.E.
Genes Dev. 5, 416-427, 1991
A:Title: Hepatocyte nuclear factor 3alpha belongs to a gene family in mammals that is ho
A:Reference number: A39533; MUID:91160974; PMID:1672118
A:Accession: A39533
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 'M', 23, 'QSDRTT', 31, 'QSDRAVAVLF', 43-44, 'T', 57-115, 'A', 117-354 <LA2>
A:Cross-references: GB:L09647
C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology
C:Keywords: DNA binding; nucleus; transcription regulation
F:119-210/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.: 2,5e-10 Length: 354
Score: 348.50 Matches: 85
Percent Similarity: 49.16% Conservative: 32
Best Local Similarity: 35.71% Mismatches: 71
Query Match: 15.92% Indels: 50
DB: 1 Gaps: 6

US-10-087-080-31 (1-1209) x S35090 (1-354)
Qy      252 GTGTGTGGCGAGGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 311
Db      82 GlyAlaGlySerGlyThrGlyGlySerAlaSerGlyTyrGlyAlaProGlyProGlyLe 101
Qy      312 CGCGAGGGGTGCACGCGACGACGATATACGGGGGGGGCGCC-----AAGCCCCC 359
Db      101 uValHisGlyLysGluMetAlaLysGlyTyrArgArgProLeuThrHisAlaLysProPr 121
Qy      360 CTACTCGTACATCGCGCTCATCGCCATCGCCATCCCGACTCCGCGCGCGCGCGCGCTTAC 419
Db      121 ofySerTyrIleSerIleThrMetAlaIleGlnAlaProGlyLysMetLeuTh 141
Qy      420 GCTGGCGGAGATCAACAGTACTCATGGCAAGTTCCCTTTTCGCGCGCGAGCTACAC 479

```

```

Db      141 rLeuSerGluIleTyrGlnTrpIleMetAspLeuPheProTyrTyrArgGluAsnGln 161
Qy      480 GGCTGGCGCAACTCGTGCAGCCACCACTTCCTCGCTCAACGACTGCTTCGTCAGGTGCT 539
Db      161 nArgTrpGlnAsnSerIleArgHisSerLeuSerPheAsnAspCysPheValVal 181
Qy      540 GCAGCGACCCCTCGCGCCCTGGGGCAAGGACAACTACTGATGCTCAACCCCAACAGCGA 599
Db      181 aArgSerProAspLysPro---GlyLysGlySerTyrTrpAlaLeuHisProSerSer 200
Qy      600 GTACACCTTCGCGAGGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 645
Db      200 yAsnMetPheGluAsnGlyCysTyrLeuArgGlnLysArgPheLysLeuGluGlu 220
Qy      645 ----- 645
Db      220 sAlaLysLysGlyAsnSerAlaThrSerAlaThrArgAsnGlyThrValGlySerAlaTh 240
Qy      646 -----AGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671
Db      240 rSerAlaThrThrAlaAlaThrAlaValThrSerProAlaGlnProGlnProThrPr 260
Qy      672 CGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716
Db      260 oProSerGluProGluAlaGlnSerGlyGluAspValGlyGlyLeuAspCysAlaSerPr 280
Qy      717 GCCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 776
Db      280 oProSer---SerAlaProTyrPheThrGlyLeuGluLeuProGly-----GluLeu 297
Qy      777 CGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
Db      297 sLeuAspAlaProTyrAsnPheAsnHisProPheSerIleAsnAsnLeuMet 314

RESULT 13
JH0672
brain factor 1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 24-Sep-1999
C:Accession: JH0672
R:Tao, W.; Lai, E.
Neuron 8, 957-966, 1992
A:Title: Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork hea
A:Reference number: JH0672; MUID:92265309; PMID:1350202
A:Accession: JH0672
A:Molecule type: mRNA
A:Residues: 1-480 <TAO>
A:Cross-references: GB:M87634; NID:G203134; PIDN:AAA40812.1; PID:G203135
A:Experimental source: brain
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
C:Keywords: DNA binding; transcription factor
F:162-271/Region: DNA binding #status predicted
F:172-263/Domain: fork head DNA-binding domain homology <FHD>

Alignment Scores:
Pred. No.: 3,53e-10 Length: 480
Score: 345.00 Matches: 127
Percent Similarity: 40.98% Conservative: 41
Best Local Similarity: 30.98% Mismatches: 141
Query Match: 15.76% Indels: 101
DB: 2 Gaps: 19

US-10-087-080-31 (1-1209) x JH0672 (1-480)
Qy      22 CTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
Db      81 ProProGlnAlaArgGlyAlaProAlaAlaAspAspLysGly----- 95
Qy      82 CGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
Db      96 ---ProGlnProLeuLeuLeuProProSerAlaAlaLeuAspGlyAlaLysAlaAspAla 114

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: March 29, 2004, 11:53:16 ; Search time 17.5 seconds
(without alignments)
7194.608 Million cell updates/sec

Title: US-10-087-080-31
Perfect score: 2189
Sequence: 1 atgaagtggagggttcgt.....tggagacgctcctagattga 1209

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO spool p/US10087080/runat 29032004 114526 6530/app query.fasta_1.1351
-DB=Swissprot 42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=prc -NOR=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087080 -CGEN 1 1 16 @runat 29032004 114526 6530 -NCPU=6 -ICPU=3
-NO MMAP -LARGSEQURY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2081.5	95.1	403	1 FXQ1 HUMAN	Q9C009 homo sapien
2	1589	72.6	400	1 FXQ1 MOUSE	O70220 mus musculus
3	1458.5	66.6	392	1 FXQ1 RAT	O63244 rattus norv
4	558	25.5	497	1 FXD2 HUMAN	O60548 homo sapien
5	543.5	24.8	440	1 FXGA CHICK	Q98937 gallus gall
6	517	23.6	465	1 FXD1 HUMAN	Q16676 homo sapien
7	505	23.1	456	1 FXD1 MOUSE	Q61345 mus musculus
8	467	21.3	478	1 FXD3 HUMAN	Q94J05 homo sapien
9	465.5	21.3	465	1 FXD3 MOUSE	O61060 mus musculus
10	445.5	20.4	388	1 FXE3 MOUSE	O90Y14 mus musculus
11	442.5	20.2	319	1 FXD3 CHICK	O13461 homo sapien
12	440.5	20.1	394	1 FXD3 HUMAN	P79772 gallus gall
13	436.5	19.9	376	1 FXS1 HUMAN	O00358 homo sapien
14	431.5	19.7	444	1 FXF2 HUMAN	O12947 homo sapien
15	430.5	19.7	553	1 FXC1 MOUSE	Q61572 mus musculus
16	429.5	19.6	553	1 FXC1 HUMAN	Q12948 homo sapien
17	415.5	19.0	376	1 FXL2 HUMAN	P58012 homo sapien
18	412	18.8	416	1 FXD5 HUMAN	Q8WXT5 homo sapien

19	407	18.6	439	1 FXD4 HUMAN	Q12950 homo sapien
20	397.5	18.2	472	1 FXH1 BRARE	Q91961 brachydanio
21	397	18.1	345	1 FXL1 HUMAN	Q12952 homo sapien
22	394.5	18.0	408	1 FXDL HUMAN	Q9NU39 homo sapien
23	394.5	18.0	444	1 FXD4 MOUSE	Q60688 mus musculus
24	387	17.7	473	1 FXD3 HUMAN	P55317 homo sapien
25	384.5	17.6	501	1 FXC2 HUMAN	Q99958 homo sapien
26	384	17.5	494	1 FXC2 MOUSE	Q61850 mus musculus
27	371.5	17.0	457	1 HN3B HUMAN	Q9Y261 homo sapien
28	370.5	16.9	468	1 HN3A MOUSE	P35582 mus musculus
29	361.5	16.5	337	1 FXL1 MOUSE	Q64731 mus musculus
30	360.5	16.5	353	1 FXF1 MOUSE	Q61080 mus musculus
31	359	16.4	354	1 FXF1 HUMAN	Q12946 homo sapien
32	356	16.3	459	1 HN3B MOUSE	P35583 mus musculus
33	356	16.3	466	1 HN3A RAT	P23512 rattus norv
34	354.5	16.2	456	1 FD3 DROME	Q02361 drosophila
35	348.5	15.9	354	1 HN3G RAT	P32183 rattus norv
36	346.5	15.8	428	1 FXB2 MOUSE	Q64733 mus musculus
37	345	15.8	480	1 FXGB RAT	Q00939 rattus norv
38	344.5	15.7	500	1 FXE2 HUMAN	O99526 homo sapien
39	344	15.7	350	1 HN3G HUMAN	P55318 homo sapien
40	343.5	15.7	469	1 FXGA HUMAN	P55316 homo sapien
41	341	15.6	481	1 FXGB MOUSE	Q60987 mus musculus
42	340	15.5	351	1 FXI1 HUMAN	Q12951 homo sapien
43	338.5	15.5	477	1 FXGB HUMAN	P55315 homo sapien
44	335.5	15.3	253	1 HN3G MOUSE	P35584 mus musculus
45	329.5	15.1	325	1 FXB1 MOUSE	Q64732 mus musculus

ALIGNMENTS

RESULT 1
FXQ1_HUMAN STANDARD; PRT; 403 AA.
ID AC Q9C009; Q9NS06;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Forkhead box protein Q1 (Hepatocyte nuclear factor 3 forkhead homolog
DE 1) (HNF-3/forkhead-like protein 1) (HNF-1).
GN FOXQ1 OR HFL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=21614676; PubMed=11747606;
RA Bieller A., Pasche B., Frank S., Glaeser B., Kunz J., Witt K.,
Zoll B.;
RT "Isolation and characterization of the human forkhead gene FOXQ1.";
RL DNA Cell Biol. 20:555-561(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21207087; PubMed=11309849;
RA Hong H.-K., Noversteke J.K., Headon D.J., Liu T., Sy M.S.,
Justice M.J., Chakravarti A.;
RT "The winged helix/forkhead transcription factor Foxq1 regulates
RT differentiation of hair in satin mice.";
RL Genesis 29:163-171(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
Diatchenko L., Marusica K., Farmer A.A., Rubin G.M., Hong L.,
Scapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywinski M.R., Skalska U., Marra M.A.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in the stomach,
 CC trachea, bladder and salivary gland.
 CC -!- SIMILARITY: Contains 1 fork-head domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL; AF225950; AAK06339.1; -;
 DR EMBL; AF153341; AAF75586.1; -;
 DR EMBL; BC053850; AAH53850.1; -;
 DR HSP; O63245; 2HPH.
 DR Genew; HGNC:20951; FOXO1.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DNA_BIND 119 214 FORK-HEAD.
 FT DONAIN 13 103 ALA/GLY-RICH.
 FT DONAIN 221 397 PRO-RICH.
 FT CONFLICT 49 52 NSP -> KPS (IN REF. 2).
 FT CONFLICT 60 61 PP -> TQ (IN REF. 2).
 FT CONFLICT 386 386 S -> L (IN REF. 2).
 FT CONFLICT 395 395 P -> S (IN REF. 2).
 SQ SEQUENCE 403 AA; 41491 MW; B552255AEAC6929B CRC64;

Alignment Scores:

Pred. No.: 1.74e-77 Length: 403
 Score: 2081.50 Matches: 396
 Percent Similarity: 98.51% Conservative: 1
 Best Local Similarity: 98.26% Mismatches: 5
 Query Match: 95.09% Indels: 1
 DB: 1 Gaps: 1

US-10-087-080-31 (1-1209) x FOXO1_HUMAN (1-403)

Qy 1 ATGAAGTTGAGGTGTTGTCCTCGCGCGCCAGCGGACAAAGCAGGCGAGTGACCTG 60
 Db 1 MetLysLeuGluValPheValProArgAlaAlaHisGlyAspLysGlnGlySerAspLeu 20
 Qy 61 GAGCGCGCGCGCGCAGCAGCGCGTCCCGCTGTCCGCGCGCGGAGACGACTCCCTG 120
 Db 21 GluGlyAlaGlyGlySerAspAlaProSerProLeuSerAlaAlaGlyAspAspSerLeu 40
 Qy 121 GGCTCAGATGGGACTCGCGGCGC---AAGCGCTCGCGCGCGCGCGCGCGCAGAGATACG 177
 Db 41 GlySerAspGlyAspCysAlaAlaAenSerProAlaAlaGlyGlyAlaArgAspPro 60
 Qy 178 CAGGCGCAGCGCAGCAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 237
 Db 61 ProGlyAspGlyGluGlnSerAlaGlyGlyProGlyAlaGluGluAlaAlaProAla 80

RESULT 2

FOXO1_MOUSE
 ID FOXO1_MOUSE PRT; 400 AA.
 AC O70220;
 DT 28-FEB-2003 (Rel. 41, Created)

Qy 238 GCAGCTCTCTCAGCGGTGGTGGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCG 297
 Db 81 AlalaAlaAlaAlaValValAlaGluGlyAlaGluAlaGlyAlaGlyProGlyAla 100
 Qy 298 GCGCGCGCGGAGCGCGCGGAGGTGTGACAGCAAGCCATATACGCGCGCGCGCGCG 357
 Db 101 GlyGlyAlaGlySerGlyGluGlyAlaArgSerLysProTyThrArgArgProLysPro 120
 Qy 358 CCCTACTCTGACATCGCGCTCATCGCGCATCGCGCATCGCGCGCGCGCGCGCGCTTG 417
 Db 121 ProTySerTyrlaAlaLeuIleAlaMetAlaIleArgAspSerAlaGlyGlyArgLeu 140
 Qy 418 ACGTGGCGGAGATCAACAGATCACTCATGGCAAGTTCCTTTTTCGCGCGCGAGTAC 477
 Db 141 ThrLeuAlaGluIleAsnGluTyLeuMetGlyLysPheProPhePheArgGlySerTy 160
 Qy 478 ACGGCTGGCGCAACTCCGTGGCGCAACCTTTCCTCAACGACTGCTTCGTCAAGGTG 537
 Db 161 ThrGlyTrpArgAsnSerValArgHisLeuSerLeuAsnAspCysPheValLysVal 180
 Qy 538 CTGGCGACCCCTCGCGCGCGCGTGGGCAAGGACAACTACTGATGCTCAACCCCAACAGC 597
 Db 181 LeuArgAspProSerArgProTyGlyLysAspAsnTyTrpMetLeuAsnProAsnSer 200
 Qy 598 GAGTACACCTTCGCCACACGGGTCTTCGCCCGCGCGCGCGCGCGCGCGCGCGCG 657
 Db 201 GluTyThrPheAlaAspGlyValPheArgArgLysArgLeuSerHisArgAla 220
 Qy 658 CCGGTCCCGCGCGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
 Db 221 ProValProAlaProGlyLeuArgProGluGluAlaProGlyLeuProAlaAlaPro 240
 Qy 718 CCGCGCGCGCGCGCGCGCGTCCGCCCATCGCTCGCGCGCGCGCGCGCGCGCGCG 777
 Db 241 ProAlaProAlaAlaProAlaSerProArgMetArgSerProAlaArgGlnGluArg 260
 Qy 778 GCAGCGCGCGCGCGCGAAGTCTCCAGCTCTTCGCGCATCGCAGCATCTCGCGCGCGCG 837
 Db 261 AlaSerProAlaGlyLysPheSerSerPheAlaIleAspSerIleLeuArgLysPro 280
 Qy 838 TTCGCGAGCGCTCGCTCAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 897
 Db 281 PheArgSerArgArgLeuArgAspThrAlaProGlyThrLeuGlnTrpGlyAlaAla 300
 Qy 898 CCTTGGCGCGCGTGGCGCGCGTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCTG 957
 Db 301 ProCysProProLeuProAlaPheProAlaLeuLeuProAlaAlaProCysArgAlaLeu 320
 Qy 958 CTGCGCTCTGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1017
 Db 321 LeuProLeuCysAlaTyrglyAlaGlyGluProAlaArgLeuGlyAlaArgGluAlaGlu 340
 Qy 1018 GTGCGACCGCGCGCGCGCGCGTCTGCTGCGACCTCTCCCGCGCGCGCGCGCGCGCGAG 1077
 Db 341 ValProThrAlaProProLeuLeuAlaProLeuProAlaAlaAlaAlaProAlaLys 360
 Qy 1078 CCACTCCGAGCA 1137
 Db 361 ProLeuArgGlyProAlaAlaGlyGlyAlaHisLeuTyCysProLeuArgLeuProAla 380
 Qy 1138 GCCTTCAGCGCGCGCTTAGTCCGCGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGAG 1197
 Db 381 AlaLeuGlnAlaAlaSerValArgArgProGlyProHisLeuProTyProValGluThr 400
 Qy 1198 CTCTAGCT 1206
 Db 401 LeuLeuAla 403

QY 1147 GCGCCCTAGTCGCGNCCTCTGGCCCGCACCTGCTGACCGGTGAGACGCTCTAGCT 1206
 Db |||||
 373 AlalaalaalaCysGlyProGlyProHisLeuSerTyrArgValGluThrLeuLeuAla 392

RESULT 4

FXD2_HUMAN STANDARD; PRT; 497 AA.
 AC O60548;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 10-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Forkhead box protein D2 (Forkhead-related protein FKHL17) (Forkhead-
 DE related transcription factor 9) (FHEAC-9).
 GN FOXD2 OR FKHL17 OR FHEAC9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98066765; PubMed=9403061;
 RA Ernster S., Betz R., Lagercrantz S., Larsson C., Erickson S.,
 RA Cederberg A., Carlsson P., Ernster S.;
 RT "Cloning and characterization of fheac-9 (FKHL17), a novel kidney-
 RT expressed human forkhead gene that maps to chromosome 1p32-p34.";
 RL Genomics 46:78-85(1997).
 RN [2]
 RP REVISIONS.
 RA Ernster S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Probable transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Kidney specific.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF042832; AAC15421.1; -.
 CC GO; GO:0003700; F:transcription factor activity; TAS.
 CC TRANSFAC; T02485; FOXD2.
 CC MIM; 602211; -.
 CC InterPro; IPR001766; TF_Fork_head.
 CC Pfam; PF00250; Fork head; 1.
 CC PRINTS; PR00053; FORKHEAD.
 CC ProDom; PD000425; TF_Fork_head; 1.
 CC SMART; SM00339; FH; 1.
 CC PROSITE; PS00657; FORK_HEAD_1; 1.
 CC PROSITE; PS00658; FORK_HEAD_2; 1.
 CC PROSITE; PS00339; FORK_HEAD_3; 1.
 CC DNA-binding; Nuclear protein; Transcription regulation.
 CC DOMAIN 90 94 POLY-ALA.
 CC FT DOMAIN 101 104 POLY-ALA.
 CC FT DNA BIND 126 217 FORK-HEAD.
 CC FT DOMAIN 247 250 POLY-ALA.
 CC FT DOMAIN 296 306 POLY-ALA.
 CC FT DOMAIN 398 409 POLY-GLY.
 CC FT DOMAIN 421 426 POLY-GLY.
 CC FT DOMAIN 442 445 POLY-ALA.
 CC SQ SEQUENCE 497 AA; 49007 MW; EAAFA98D216BE019 CRC64;

Alignment Scores:

Pred. No.: 3,328-16 Length: 497
 Score: 558.00 Matches: 176
 Percent Similarity: 44.75% Conservative: 37

Best Local Similarity: 36.97% Mismatches: 152
 Query Match: 25.49% Indels: 111
 DB: Gaps: 21

US-10-087-080-31 (1-1209) x FXD2_HUMAN (1-497)

QY 37 GGGGCAAGCAGGGCAGTGCCTGGAGGGCGGGCGGCGAGCGAGCGCGCGTCCCGGTG 96
 Db |||||
 32 GlyGlySerGlyGlyGlyLeuProAlaAsgSerGlyProArgAlaProArgAspVal 51
 QY 97 TCGGGCGGGGAGAGCAGCTCCCTGGCTCAGATGGGGACTCGCGCGCAAGCCGTCGCG 156
 Db |||||
 52 LeuProHisGlyHisGlu-----ProAla 60
 QY 157 GGGCGGGCGCCAGAGATACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207
 Db |||||
 61 GluGluAlaGluAlaAspLeuAlaGluAspGluGluGluSerGlyGlyCysSerAspGly 80
 QY 208 GGGCGGGCGGGAGGAGCGGATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 267
 Db |||||
 81 GluProArgAlaLeuAlaSerArgGlyAlaAlaAlaAlaAlaGlySerProGlyProGly 100
 QY 268 GGGGAGGCC-----GGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 321
 Db |||||
 101 AlaAlaAlaAlaArgGlyAlaAlaGlyProGlyProGlyProProSerGlyGlyAlaAla 120
 QY 322 GCACGCGAGCAAGCCATATACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 381
 Db |||||
 121 ThrArgSer---ProLeuVal-----LysProProTyrSerTyrAlaLeuLeu 136
 QY 382 GCATGGCCATCCGCGACTCGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 441
 Db |||||
 137 ThrMetAlaLeuGlnSerProLysLysArgLeuThrLeuSerGluIleCysGluPhe 156
 QY 442 CTCATGGGCAAGTTCCTCTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 501
 Db |||||
 157 IleSerGlyArgPheProTyrTyrArgGluLysPheProAlaTrpGlnAsnSerIleArg 176
 QY 502 CACACCTTTTCCTCAACGACTGTCTCAAGGTGTGCGCGACCCCTCGCGGCCCTCGG 561
 Db |||||
 177 HisAsnLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro--- 195
 QY 562 GCACAGGCAACTACTGATGCTCAACCCCAACAGCGAGTACACTTCGCGCGCGGGGTC 621
 Db |||||
 196 GlyLysGlyAsnTyrTrpThrLeuAspProLysAlaAspMetPheAspAsnGlySer 215
 QY 622 TTCGCGCGCGCGCGCGCGCTACGCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 681
 Db |||||
 216 PheLeuArgArgLysArgPheLys---ArgGlnProLeuProProProHisProHis 234
 QY 682 CCGGAGGAGGCGCGCG----- 696
 Db |||||
 235 ProHisProHisProGluLeuLeuArgGlyAlaAlaAlaAlaAlaGlyAspProGly 254
 QY 697 -----GGGCTC 702
 Db |||||
 255 AlaPheLeuProGlyPheAlaAlaTyrGlyAlaTyrGlyTyrGlyTyrGlyLeuAlaLeu 274
 QY 703 CCGCGC-----GCCCG 745
 Db |||||
 275 ProAlaTyrGlyAlaProProProGlyProAla-ProHisProHisProHisProHisPr 294
 QY 746 -----GCATGGCTCG 783
 Db |||||
 294 ohisAlaPheAlaPheAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 314
 QY 784 CCGCGGGCAAGTTCCTCCAGCTCTTCGCCATCCACAGCATCTCTGC----- 829
 Db |||||
 314 OdlyArgAlaAlaAlaProProProGlyProProThrAlaSerValPheAlaGlyAlaGl 334
 QY 830 -GCAAGCCCTTCGCGCGCGCTCGCTCAGGACACGGCGCC-----CCGGA 874
 Db |||||
 334 yserAlaProAlaProAlaProAlaSerGlySerGlyProGlyProGlyProAlaGlyLe 354

QY 875 -CGAGCTTTCAGTGGGGCGCGCCCTGCTCCCGCGCTTCCCGCGCT---- 929
 Db 354 uProAlaLeuGlyAlaGluLeuGlyCys-AlaLysAlaPheTyrProAlaSerLeuS 374
 QY 930 -----CTTCCCGCGCGCCCTGAGGCG---CTGCTGGCTTCCGCG-----GTAC 975
 Db 374 eProProAlaAlaGlyThrAlaGlyLeuProThrAlaLeuLeuArgGlnGlyLeuL 394
 QY 976 GCGCGGGCGAGCGCGCGCTGGCGCGCGAGCGCGAGGTGCCACCGACCGC--- 1031
 Db 394 yThrAspAlaGlyGlyGlyAlaGly---GlyGlyGlyAlaGlyAlaGlyGlnArgProS 413
 QY 1032 -----GCGCGCCCTCTGCTTGC- 1049
 Db 413 ePheSerIleAspHisIleMetGlyHisGlyGlyGlyAlaAlaProProGlyAlaG 433
 QY 1050 -----ACCTCTCCCGCGCGCGCGCCCGCCAGCACTCCGAGCGCGCGCGCGCGC 1104
 Db 433 lYgluGlySerPro---GlyProPheAlaAlaAlaGlyProGlyGlyGlnAlaG 452
 QY 1105 GCGCACCTGTACTGCGCCCTCGGCTGCGCGCGCGCTGAGCGCGCTAGTCCGNCCT 1164
 Db 452 inValLeuAlaMetLeuThrAlaProAlaLeuAlaProValAlaGly-HisIleArg--- 470
 QY 1165 CTGCGCGCGCACCTGTCTGCTACCGGTGGAGCGCTCTAGCT 1206
 Db 471 -----LeuSerHisProGlyAspAlaLeuLeuSer 480

RESULT 5
 FXGA_CHICK STANDARD; PRT; 440 AA.
 ID FXGA_CHICK
 AC Q9837;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Forkhead box protein G1A (forkhead-related protein FKHL2)
 DE (transcription factor BF-2) (Brain factor 2) (BF2) (CBF-2) (T-14-6).
 GN FOXG1A OR FKHL2 OR HFHF2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Retina;
 RX MEDLINE=96338226; PubMed=8757134;
 RA Yuasa J., Hirano S., Yamagata M., Noda M.;
 RT "Visual projection map specified by topographic expression of
 RT transcription factors in the retina.";
 RL Nature 382:632-635(1996).
 CC -!- FUNCTION: May determine the nasotemporal axis of the retina, and
 CC consequently specify the topographical projection of the retinal
 CC ganglion-cell axons to the tectum by controlling expression of
 CC their target genes.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- TISSUE SPECIFICITY: Retina.
 CC -!- DEVELOPMENTAL STAGE: Can be detected in regions including
 CC primordial retina and neuroepithelium by embryonic day 2 (E2). At
 CC E3, expressed in the temporal retina and associated pigment
 CC epithelium as well as in part of the diencephalon, and at E7 is
 CC expressed in retinal ganglion cells. Levels begin to decline from
 CC E4 and almost disappear by E10.
 CC -!- SIMILARITY: Contains 1 fork-head domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; U47276; AAB08467.1; --
 DR PIR; S71795; S71795.
 DR HSSP; O63245; 2HFH.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00339; FORK_HEAD_3; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein;
 KW Developmental protein.
 FT DOMAIN 39 43 POLY-ARG.
 FT DOMAIN 49 58 POLY-GLU.
 FT DOMAIN 77 88 POLY-GLY.
 FT DOMAIN 109 112 POLY-ALA.
 FT DOMAIN 113 137 POLY-GLY.
 FT DNA_BIND 143 237 FORK-HEAD.
 FT DOMAIN 321 330 POLY-PRO.
 SQ SEQUENCE 440 AA; 44669 MW; 82266C2E1E103A48 CRC64;
 Alignment Scores:
 Pred. No.: 1.3e-15 Length: 440
 Score: 543.50 Matches: 164
 Percent Similarity: 48.63% Conservative: 31
 Best Local Similarity: 40.90% Mismatches: 142
 Query Match: 24.83% Indels: 65
 DB: 1 Gaps: 19

US-10-087-080-31 (1-1209) x FXGA_CHICK (1-440)

QY 103 GCGGAGACACTCCCTGGGCTCAGATGGGAGTGGCGGCGCAAGCG---TCCGCGGCG 159
 Db 60 AlaGlyAsp-----LeuHisAspAspAlaLeuLeuProArgSerProValArgAlaGly 77
 QY 160 GCGGCGCCAGACATACGAGCGCGCGCGAGCGAGTGGCGGAGCGCGCGCGCG 219
 Db 78 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 97
 QY 220 GAGGA-----GGCATCCCGCAGCAGCTGCTGCAGCGGTGGTGGCGAGGG 266
 Db 98 SerArgProProSerArgGlyGlyProGlnLysAlaAlaAlaGlyGlyGlyAla 117
 QY 267 CGCGAGCG 326
 Db 118 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 137
 QY 327 CAGCAAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
 Db 137 YLysAsnSerLeuVal-----LysProProTyrSerTyrIleAlaLeuIleThrMe 154
 QY 387 GGCATCCGCGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 446
 Db 154 LAlaLeuLysSerProLysArgLeuThrLeuSerGluIleCysGluPheIleSe 174
 QY 447 GGGCAAGTTCCCTTTTCCGCGGCGAGTACAGCGGCTGGCGCAACTCGTCCGCCA 506
 Db 174 rGlyArgPheProTyrTyrArgGluLysPheProAlaLysPheProAlaLysPhe 194
 QY 507 CTTTTCCTCAACGACTGCTTCTCAAGGTGCTGGCGCGCGCGCGCGCGCGCGCG 566
 Db 194 nLeuSerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro--Gly 213
 QY 567 GGAACAATCTGATGCTCAACCCCAAGCAGGAGTACACTTCGCGCGCGGCTTCG 626
 Db 213 sGlyAsnTyrTrpThrLeuAspProGluSerAlaAspMetPheAspAnglySerPhe 233
 QY 627 CCGCGCGCGCGCAAGCGCTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
 Db 233 uArgArgArgLysArgPheLys---ArgGlnGlnLeuProAlaProGluLeuLeuAr 252

QY 687 GGAGGCGCGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT-----C 740
 Db 252 SALavalasp-----ProAlaAla-PheLeuProGlnProProGlnProProGlnG 270
 QY 741 GCCCGCAGCGCT-----C 764
 Db 270 InProProCysAlaTyrGlyProTyrGlyCysGlyTyrGlyLeuGlnProTyrH 290
 QY 765 CCAGGAGGAGCGG-----CCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 806
 Db 290 isProHisSerAlaLeuPheAlaPheHisPheSerProProProArgGlnProPro 310
 QY 807 CTTCCGATGACAGCATCTCGCGAGCCCTTCGCGAGCCCTCGCGCTCGCGCTA 866
 Db 310 laAlaProAlaGlyAla-ProAlaAlaLeuProProProProProProProPro 329
 QY 867 CCGCGGAGCAGCGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 914
 Db 330 ProArg-----ArgArgAlaProLeuLeuProAlaAlaGluLeuAla 343
 QY 915 CGCGTTCCCGCGCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 974
 Db 344 ArgThrProPheGlyTyrProHisProLeu-----GlyProAlaLeuAlaSerLeu 361
 QY 975 CGCT 1034
 Db 362 HisAlaAlaLysProGlySer-----GlyAlaAlaValAlaArg 374
 QY 1035 GCCCTCTCTGTCACCTCT---CCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1082
 Db 375 SerProPheSerIleGluSerIleIleGlyGlyProGlyProGlyLeuGlyAlaGly 394
 QY 1083 CCGAGCT 1133
 Db 395 ProAlaProGlyAlaGlySerCysAlaSerGlnSerGlyAlaAlaThrGlyLeuSer 414
 QY 1134 CGCAGCGCTGCGAGCGCGCTTGTAGTCGCGCTCTCGCGCGCGCGCGCGCT 1190
 Db 415 ArgSerLeuGlySerGlyLeuAlaProAlaAla---AlaLeuProAlaAlaProGly 432

RESULT 6
 FXDI_HUMAN STANDARD; PRT; 465 AA.
 AC Q16676; Q12949;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Forkhead box protein d1 (Forkhead-related protein FKHL8) (Forkhead-
 related transcription factor 4) (FREAC-4).
 GN FOXD1 OR FKHL8 OR FREAC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96355467; PubMed=7957066;
 RA Ennasson P., Eriou S., Hulander M., Cederberg A., Hellqvist M.,
 RA Carlsson P., Eriou S.,
 RT "Characterization of the human forkhead gene FREAC-4. Evidence for
 RT regulation by Wilms' tumor suppressor gene (WT-1) and p53."
 RL J. Biol. Chem. 271:21094-21099(1996).
 RN [2]
 RP SEQUENCE OF 120-225 FROM N.A.
 RX MEDLINE=95045392; PubMed=7957066;
 RA Eriou S., Hellqvist M., Samuelsson L., Ennasson P., Carlsson P.,
 RT "Cloning and characterization of seven human forkhead proteins:
 RT binding site specificity and DNA binding."
 RL EMBO J. 13:5002-5012(1994).
 CC -/- FUNCTION: Binding of FREAC-3 and FREAC-4 to their cognate sites
 CC results in bending of the DNA at an angle of 80-90 degrees.
 CC -/- SUBCELLULAR LOCATION: Nuclear.
 CC -/- SIMILARITY: Contains 1 fork-head domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC ENBL; US9832; AAC50661.1; -
 CC ENBL; US9831; AAC50660.1; -
 CC ENBL; U13222; AAA92039.1; -
 CC PIR; G02738; G02738.
 CC PIR; S51627; S51627.
 CC HSP; O63245; 2HPH.
 CC TRANSFAC; T02472; -
 CC Genew; HGNC; 3802; FOXD1.
 CC MIM; 601091; -
 CC GO; GO:0003700; P:transcription factor activity; TAS.
 CC InterPro; IPR001766; TF_Fork_head.
 CC Pfam; PF00250; Fork_head; 1.
 CC PRINTS; PR00053; FORKHEAD.
 CC ProDom; PD000425; TF_Fork_head; 1.
 CC SMART; SM00339; FH; 1.
 CC PROSITE; PS00657; FORK_HEAD_1; 1.
 CC PROSITE; PS00658; FORK_HEAD_2; 1.
 CC PROSITE; PS00659; FORK_HEAD_3; 1.
 CC Transcription regulation; DNA-binding; Nuclear protein.
 KW DOMAIN 26 34 POLY-GLU.
 FT DOMAIN 39 43 POLY-GLY.
 FT DOMAIN 52 57 POLY-ARG.
 FT DOMAIN 69 72 POLY-GLU.
 FT DOMAIN 73 76 POLY-ASP.
 FT DOMAIN 97 113 POLY-GLY.
 FT DNA_BIND 124 215 FORK-HEAD.
 FT DOMAIN 231 234 POLY-ALA.
 FT DOMAIN 252 256 POLY-ALA.
 FT DOMAIN 259 266 POLY-PRO.
 FT DOMAIN 293 303 POLY-ALA.
 FT DOMAIN 309 315 POLY-PRO.
 FT DOMAIN 375 390 POLY-ALA.
 FT DOMAIN 428 434 POLY-ALA.
 SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBF AE CRC64;
 Alignment Scores:
 Pred. No.: 1.5e-14 Length: 465
 Score: 517.00 Matches: 157
 Percent Similarity: 44.21% Conservative: 34
 Best Local Similarity: 36.34% Mismatches: 132
 Query Match: 23.62% Indels: 109
 DB: 1 Gaps: 16
 US-10-087-080-31 (1-1209) x FXDI_HUMAN (1-465)
 QY 60 GGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCCT 119
 Db 39 GlyGlyGlyGlyGlyGlyProArgLeuAlaValPro----- 49
 QY 120 GGGCTCAGATGGGGAGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
 Db 50 -----AlaGlnArgArgArgArgArgArgArgArgArgArg 60
 QY 180 GGGCGCAGCG 203
 Db 61 GlyGluAspGluLeuGluAspLeuGluGluGluGluGluGluLeuAla 80
 QY 204 AGCG 263
 Db 81 ProProAlaGlyGlySerProAlaProGlyProAlaProAlaGlyAla 100
 QY 264 GGGCG 323
 Db 101 GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 119

QY 324 ACCGACAGCCATATACGGCGCGGCCAAGCCGCCCTACTCTGATACATCGCCCTCATGCC 383
Db 119 alyAsn---ProLeuVal-----LysProProTyrSerTyrIleAlaLeuIlePh 135
QY 384 CATGGCCATCGGACTCGCGCGCGCGCTGACGCTGGCGGAGATCAACGAGTACCT 443
Db 135 rMetAlaIleLeuGlnSerProlyslsYargLeuThrLeuSerGluileCysGluPheI 155
QY 444 CATGGCAGAGTCCCTCTTTTCGCGCGGACATACAGCGGCTGCGCACTCGTGGCGCA 503
Db 155 eSerGlyArgPheProTyrTyrArgGluLysPheProAlaTyrGlnAsnSerIleArgH 175
QY 504 CAACCTTTCGCTCAACGAGCTCTTCTCAAGGTCTGCGCGACCCCTCGCGCCCTGGGG 563
Db 175 sAnLeuSerLeuAsnAspCysPheValylsIleProA-gLuproGlyAsnPro---Gl 194
QY 564 CAAGGACAATCTAGTATGCTCAACCCCAACAGCAGTACACTTTCGCGCGAGGGTCTT 623
Db 194 yLysGlyAsnTyrTrpThrLeuAspProGluSerAlaAspMetPheAsnGlySerPh 214
QY 624 CCGCGCGCGCGAGCGCTCAGCCAGCGCGCGCGCTCCCGCGCGCGCTCGCGGC 683
Db 214 eLeuArgArgGlyArgPheLys---ArgGlnProLeuLeuProProAsnAlaAla 233
QY 684 CGAGGAGGCC-----CCGCGCTCCCGCGCGC 710
Db 233 aAlaGluSerLeuLeuArgGlyAlaGlyAlaAlaGlyGlyAlaGlyAspProAlaAl 253
QY 711 CCGG-----CCGCGCGCGCGCGCGCGCGCGCTCGC----- 742
Db 253 aAlaAlaLeuPheProProAla-ProProProProProHisAlaTyrGlyTyGlyP 273
QY 743 -----CCGATCGCTCGCTCCCGCGCGCGCGCGCGCGAGGA 773
Db 273 roTyGlyCysGlyTyGlyLeuGlnLeuProTyAlaProProSerAlaLeuPheA 293
QY 774 CGCGCGCAGCGCGCGCGCGAGTCTCCAGCTCTCGCATCGACGATCTCGCGCA 833
Db 293 laAlaAlaAlaAlaAlaAlaAlaAlaAlaPheHisProHis-----S 308
QY 834 GCGCTTCGCGCGCGCTCGCTCAGGACAGCGCGCGCGCGCGCTTCAGTGGCGCG 893
Db 308 erPro-ProProProProPro---HisGly----- 317
QY 894 CGCGCGCTCGCGCGCTCGCGCGCTTCCTCCCGCGCT-----CCT 932
Db 318 -----AlaAlaAlaGluLeuAlaAlaArgThrAlaPheGlyTyArgProHisPro 333
QY 933 CCGCGCGCGCGCTCGCGCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCG 992
Db 334 LeuGlyAlaAlaLeuProGlyProLeuProAlaSerAlaAlaGlyGlyProGly 353
QY 993 CGCGCTCGCGCGCGCGCGCGCG-----CGAGTGCACCGCAC 1028
Db 354 AlaserAlaLeuAlaArgSerProPheSerIleGluSerIleIleGlySerLeuGly 373
QY 1029 CGCGCGCGCTCTCTGTGACCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGAG 1088
Db 374 ProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 393
QY 1089 CCG 1148
Db 394 Pro-----SerProValAlaAlaProAlaProGlySerSerGlyGly 408
QY 1149 GGCCTAGTCGNGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1178
Db 409 GlyCysAlaAlaGlnAlaAlaAlaValGlyPro 418
RESULT 7
ID FXD1_MOUSE
AC Q61345; PRT; 456 AA.

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-
related transcription factor 4) (FREAC-4).
GN FOXD1 OR FKHL8 OR FREAC4 OR FHBF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Embryo;
RX MEDLINE=95114592; PubMed=7815060;
RA Hatini V., Tao W., Lai E.;
RT "Expression of winged helix genes, BF-1 and BF-2, define adjacent
domains within the developing forebrain and retina."
RL J. Neurobiol. 25:1293-1309(1994).
CC -I- FUNCTION: May function in specifying positional identity in the
developing retina as well as the subdivision of the forebrain
neuroepithelium.
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -I- TISSUE SPECIFICITY: Predominantly expressed in the CNS and
temporal half of the retina. Also expressed in the condensed head
mesenchyme, metanephric blastema of the developing kidney, cortex
of the adrenal gland, condensed mesenchyme at the base of the
follicles of vibrissae and cartilage perichondrium of the
developing vertebrae.
CC -I- DEVELOPMENTAL STAGE: At E9.5 embryos, expressed in a limited
region of the neuroepithelium and also in the temporal half of the
primary optic cup and the optic stalk. At E10.5, seen in the
hypothalamus, temporal half of the optic stalk, and temporal
hemirhina. At E12.5 and E13.5 a high expression is seen in
regions of condensed mesenchyme of the head, and as
neuroepithelial cells begin to differentiate and migrate outward
from the ventricular zone, expression declines markedly. By E16.5
levels are diminished and restricted to unused pockets along the
exhausted ventricular zone.
CC -I- SIMILARITY: Contains 1 fork-head domain.
CC -I- CAUTION: Was originally (Ref.1) assigned to be BF-2 (FOXG1A).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; L38607; AAC42042.1; -
HSSP; Q63245; 2HFH.
TRANSFAC; T02293; -
MGD; MGI:1347463; Foxd1.
InterPro; IPR001766; TF_Fork_head.
Pfam; PF00250; Fork_head; 1.
PRINTS; PRO0053; FORKHEAD.
ProDom; PD000425; TF_Fork_head; 1.
SMART; SM00339; FH; 1.
PROSITE; PS00657; FORK_HEAD_1; 1.
PROSITE; PS00658; FORK_HEAD_2; 1.
PROSITE; PS50039; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
Developmental protein.
FT DOMAIN 29 32
FT POLY-GLU.
FT DOMAIN 33 36
FT POLY-ASP.
FT DOMAIN 38 46
FT POLY-GLY.
FT DOMAIN 55 58
FT POLY-ARG.
FT DOMAIN 73 76
FT POLY-ASP.
FT POLY-HEAD.
FT DNA_BIND 130 224
FT DOMAIN 261 267
FT POLY-PRO.
FT DOMAIN 293 302
FT POLY-ALA.
FT DOMAIN 308 319
FT POLY-PRO.
FT DOMAIN 395 400
FT POLY-GLY.

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Forkhead box protein D3 (HNF3/FH transcription factor genesis)
 DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HFN-2).
 GN FOXD3 OR FHF2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.,
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Labosky P.A., Kaestner K.H.;
 RT "The winged helix transcription factor Hfh2 is expressed in neural
 RL crest and spinal cord during mouse development."
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Probable transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U01047; AAA87569.1; -.
 DR EMBL; AF067421; AAC28352.1; -.
 DR HSSP; OG3245; 2HPH.
 DR TRANSFAC; T04166; -.
 DR MGD; MGI:1347473; Foxd3.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00339; FORK_HEAD_3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DOMAIN 106 115
 FT POLY-GLY.
 FT DNA_BIND 131 225
 FT FORK-HEAD.
 FT DOMAIN 252 257
 FT POLY-ALA.
 FT DOMAIN 285 270
 FT POLY-ALA.
 FT DOMAIN 275 281
 FT POLY-ALA.
 FT DOMAIN 380 399
 FT POLY-GLY.
 FT DOMAIN 447 457
 FT POLY-ALA.
 SQ SEQUENCE 465 AA; 47092 MW; 6F8B5BB3D8C7564D CRC64;

Alignment Scores:

Pred. No.:	1,778-12	Length:	465
Score:	465.50	Matches:	154
Percent Similarity:	39.17%	Conservative:	34
Best Local Similarity:	32.08%	Mismatches:	173
Query Match:	21.27%	Indels:	119
DB:	1	Gaps:	16

US-10-087-080-31 (1-1209) x FOXD3_MOUSE (1-465)

QY	37	GGGACAACAGCGGAGTACTGAGGCG	-----GCGGGCGGAGCGAGCGCGG	87
DB	5	GlySerGlySerAlaSerAspMetSerGlyGlnThrValLeuThrAlaGluAspValAsp	24	
QY	88	TCCCGCTCTCGCGCGGAGACACCTCCG	-----GGCTCAGATGGGACTGC	138
DB	25	IleAspValValGlyGluGlyAspGlyLeuGluGlyAspSerAspAlaGlyCys	44	

QY	139	-----GCGGCCAAGCGG	-----	150
DB	45	AspSerProAlaGlyProProAspLeuArgLeuAspGluAlaAspGluGlyProProVal	64	
QY	151	TCCGGCGGCGGCGGCCAGAGATAGCAGGCGACGCGGCAACAGAGTGGGAGCGGG	210	
DB	65	SerAlaHisGlyGlnSerGlnProGlnAlaLeuAlaLeuProThrGluAlaThrGly	84	
QY	211	CCGGCGCGGAGGAGCGATCCGGCAGCAGCTGTGCGAGCGGTGCTGGCGGAGGCGGG	270	
DB	85	ProGlyAsnAspThrGlyAlaProGluAlaAspGlyCysLys	103	
QY	271	GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	330	
DB	104	ValThrGlyGlyGlyProGlyAlaGlySerGlyAlaThrGlyGlyLeuThrProAsn	123	
QY	331	AAGCATATACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	390	
DB	124	LysProGlyAsnSerLeuValLysProProTySerTyIleAlaLeuIleThrMetAla	143	
QY	391	ATCCGCGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	450	
DB	144	IleLeuGlnSerProGlnLysLysLeuThrLeuSerGlyIleCysGluPheIleSerAsn	163	
QY	451	AAGTTCCTCTTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	510	
DB	164	ArgPheProTyTyArgGluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeu	183	
QY	511	TCGCTCAAGACGCTTCGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	570	
DB	184	SerLeuAsnAspCysPheValLysIleProArgGluProGlyAsnPro	202	
QY	571	AAGTTCGATGCTCAACCCCAACAGCGAGTACACTTCGCGGCGGCGGCGGCGGCGGCGG	630	
DB	203	AsnTyTrpThrLeuAspProGlnSerGluAspMetPheAspAsnGlySerPheLeuArg	222	
QY	631	CGCGGCAAGCGC	642	
DB	223	ArgArgLysArgPheLysArgHisGlnGlnGluHisLeuArgGluGlnThrAlaLeuMet	242	
QY	643	-----CTCAGCCACCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCG	678	
DB	243	MetGlnSerPheGlyAlaTySerLeuAlaAlaAlaGlyAlaGlyProTyGlyLeu	262	
QY	679	CGGCGCGGAGGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	738	
DB	263	HisProAlaAlaAlaAlaGly	280	
QY	739	TCGCGCGCGCATGCGCTCGCGCG	765	
DB	281	AlaGlyCysCysGlyAlaProValProValArgAlaThrThrArgGlySerArgAla	300	
QY	766	CAGGAGAGCGCGCGCGCG	807	
DB	300	aAlaArgSerAlaProAlaAlaLeuGlyArgAlaGlyProGlnSerGlyArgLeuArg	320	
QY	808	TTCCGCCATCGACAGCATCTCGGCAAGCCCTTCGCGAGCGG	855	
DB	320	euAlaAlaArgSerGluProAlaAlaAlaThrAlaGlnHisProGlyArgArgArgGlySerA	340	
QY	856	AGGACACGCGCGCGCGCGCGCGCTTCAGTGGGCGGCG	909	
DB	340	rgGlyHisGlyGlyArgGlyGlyHisValAlaHisGlnValArgAlaGlnCysAlaA	360	
QY	910	CTCGCGCGCTTCGCGCGCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	969	
DB	360	laValValGlnHisArgGluHisHisArgGlyIleArgAlaProGlyGlySer	378	
QY	970	GCGTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1029	
DB	379	-----AlaGlyGlyGlyGlySerGlyGlyAlaGlyGlyGlyGlyGlyGlyGlyGly	393	
QY	1030	GCGCGCGCGCTCTGCTTGCACCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1089	


```
QY 829 CGAAGCCCTTCGCGACCCCTCCTCAGGACACAGCCCGCCGCGGACGAGCTTCAGTGG 888
Db 224 -----AsnLeuGlnProGluLeuAlaGlyLeu----- 232
QY 889 GGGCGCGCGCCCTGCGCGCGCTGCGCGCTTCCCGCGCTCTCCCGCGGCGCCCTGC 948
Db 233 ---GlyAlaProGluProCysCysAlaAlaProAspAlaAlaAlaAlaAlaAla 249
QY 949 AGGCGCCCTGCTGCGCTCTCCTGCGTACGGCGCGGCGAGCGCGCGCTGGCGCGCGC 1008
Db 250 -----PheProCysAlaAlaAlaAlaAlaSerProLeu----- 261
QY 1009 GAGCGCAGGTGCACCGACCGCGCGCCCTCTCTGCTTGCACCTCTCCGCGGCGCGCC 1068
Db 262 TyrSerGlnValPro--AspArgLeuValLeuProAla---ThrArgProGlyProGlyP 280
QY 1069 CCGGCCAAGCACTCCGAGGCGCGCGCGCGCGCGCGCTGACTGCCCCCTCGCG 1128
Db 280 ro-----LeuProAlaGluProLeuLeuA 288
QY 1129 CTGCGCGCAGCCCTGCAGCGCGCTTAGTCCGNCGTCC 1166
Db 288 laLeuAlaGlyProAlaAlaAlaLeuGlyProLeuSer 300

RESULT 12
FXD3_CHICK STANDARD; PRT; 394 AA.
AC P79772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein D3 (HNF3/FH transcription factor gene) (Winged
DE helix protein CW3-3).
GN FOXD3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=97141794; PubMed=8988052;
RT Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;
RT "Aberkant cell growth induced by avian winged helix proteins.";
RL Cancer Res. 57:123-129(1997).
CC -!- FUNCTION: Probable transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U37274; AAC60066.1; -.
CC HSSP; Q63245; 2HPH.
DR TRANSFAC; T02495; -.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation.
FT DOMAIN 67 70 POLY-ALA.
FT DOMAIN 80 91 POLY-GLY.
FT DOMAIN 100 106 POLY-ALA.
```

```
FT DNA BIND 117 211 FOK-HEAD.
SQ SEQUENCE 394 AA; 40995 MW; 324A4B36B9E31899 CRC64;

Alignment Scores:
Pred. No.: 1,84e-11 Length: 394
Score: 440.50 Matches: 147
Percent Similarity: 41.52% Conservative: 39
Best Local Similarity: 32.81% Mismatches: 122
Query Match: 20.12% Indels: 140
DB: 1 Gaps: 19

US-10-087-080-31 (1-1209) x FXD3_CHICK (1-394)
QY 49 GCGAGTGCCTGGAGGCG-----GCGGCGCGACGAGCGCGCTCCCGCTCGTCG 99
Db 7 GlySerAspMetSerGlyGlnThrAlaLeuAlaAlaGluAspValAspVal 26
QY 100 GCGGCGGAGACGACTCCCTGGCTCAGATGGGAGCTGCGCGGCGCAAGCCCTCCGCGGCG 159
Db 27 GlyGluGlyAspAlaProGlyLysAspGlyGluAlaArg-SerProAlaAl 46
QY 160 G-----GCGGCGCGACGATACGAGGCGAC 186
Db 46 aLeuProLeuProLeuAspGluAlaAlaGluProGlyGluProGlu---ArgAlaAla-A 65
QY 187 GCGCAACAGAGTGGCGGAGCGCGCGCGCGAGGAGGCGATCCCGCGACGAGCTGCT 246
Db 65 rGArgAlaAlaAlaAlaArgGlnProGlyArgProGluGlyArg----- 82
QY 247 GCAGCGTGTGGCGGAGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 83 --GlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 101
QY 307 GCGGCGCGGAGGCTGCACGAGCAAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 102 AlaAlaAlaAlaAlaGlyGlnSerLysProLysSerSerLeuValLysProLysSer 121
QY 367 TACATCGCGCTCATCGCCATCGCGACTCGCGACTCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 122 TyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProGlnLysLeuThrLeuSer 141
QY 427 GAGATCAACGAGTACTCATCGGCAAGTTCCCTTTTCCCGCGCGAGTACACGCGCTGG 486
Db 142 GlyIleCysGluPheIleSerAsnArgPheProTyrTyrArgGluLysPheProAlaTrp 161
QY 487 CGCAACTCCGTCGCGCCAACTTCGCTCAACGACTCTTCGCTCAAGTCTGCTCAAGTCTG 546
Db 162 GlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPheValLysIleProArgGlu 181
QY 547 CCTCGCGCGCTGGCGCAAGCAACTACTGATGCTCAACCCCAACAGCGAGTACACC 606
Db 182 ProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAspProGlnSerGluAspMet 200
QY 607 TTCGCGCGCGGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db 201 PheAspAsnGlySerPheLeuArgArgLysArgLysArgHisGlnGlnGluHis 220
QY 661 -----GTCCCGCGCGCG 672
Db 221 LeuArgAspGlnThrAlaLeuMetMetGlnGlyPheGlyAlaTyrGlyLeuAlaGlyPro 240
QY 673 GGGCTCGCGCGCGCGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
Db 241 TyrGlyArgProTyr-----GlyLeuProProGlyAlaTyrProHisProAlaAla 257
QY 733 CCGGCGCTGCCCGCGCATGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 792
Db 257 ----- 257
QY 793 AAGTTCTCCAGCTCCTTCGCCATCGACAGATCTTCGCAAGCCCTTCGCGAGCCGCTCG 852
Db 257 ----- 257
```

QY 853 CTCAGGACAGCGCCCGGACGAGCTTCAGTGGGCGCGCGCTGCGCCGCGCTG 912
 Db 258 -----LeuGlnTyrProTyrIlePro----- 264
 QY 913 CCQCGGTTCCCGGCGCTCTCCCGCGCGCCCTGAGGCGCCCTGCGCTGCGCG 972
 Db 265 ProValGlyProMetLeuProProAlaCysPro-----LeuLeuPro----- 278
 QY 973 TACGCGGCGGCGCGCGCGCGCTGCGCGCGCGGAGGCGCGAGGTCACACGCGCG 1032
 Db 279 -----SerGlyGluLeuSerArg--LysAlaPheAenAlaGlnLeuGlyProSerLeu 295
 QY 1033 CGCGCCCTCTGCTTGCACCTCTCCCGCGCGCGCC-----CCC 1071
 Db 296 ---GlnLeuGlnLeuSerSerLeuGlyAlaAlaGlySerIleValLysSerGluProSer 314
 QY 1072 GCCAAGCCA-----CTCCGAGCGCGCGCGCGCGCGCGCGCGCGCAC 1110
 Db 315 SerArgProSerPheSerIleGluAenIleIleGlyProAlaAlaSerAla--- 333
 QY 1111 CTGTACTGCGCC-----CTGCGGCTGCGCGCGCGCGCGCGCGCGCGCTTA 1155
 Db 334 -----ProSerAlaGlnThrPheLeuArgProProValThrValGlnSerGlyLeu 350
 QY 1156 GTC---CGNCGTCT-----GGCGCGCGCGCGCGCGCGCGCGCGCTCG 1182
 Db 351 ValAlaHisGlnProLeuAlaLeuAlaArgThrThrAlaAlaIleAlaProIleLeuSer 370
 QY 1183 TACCGCGGTGGACAGCTC 1200
 Db 371 ValProThrAsnIle 376
 RESULT 13
 ID FXEL HUMAN STANDARD; PRT; 376 AA.
 AC Q00358; Q75765;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Forkhead box protein E1 (Forkhead-related protein FKHL15) (Thyroid
 DE transcription factor 2) (TFP-2).
 GN FOXE1 OR FKHL15 OR TIF2 OR TTF2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN [2]
 RN [3]
 RN [4]
 RA Maccchia P.E., Mattei M.-G., Lapi P., Fenzi G., Di Lauro R.;
 RA "Cloning, chromosomal localization and identification of two
 RA polymorphisms in the human thyroid transcription factor 2 gene
 RA (TTF2).";
 RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RX MEDLINE=98361170; PubMed=9697705;
 RA Clifton-Bligh R.J., Wentworth J.M., Heinz P., Crisp M.S., John R.,
 RA Lazarus J.H., Ludegate M., Chatterjee V.K.;
 RA "Mutation of the gene encoding human TTF-2 associated with thyroid
 RA agenesis, cleft palate and choanal atresia";
 RN Nat. Genet. 19:399-401(1998).
 RN [4]
 RP VARIANT BAMFORTH-LAZARUS SYNDROME ASN-57.
 RP MEDLINE=22155115; PubMed=12165566;
 RA Castanet M., Park S.M., Smith A., Bost M., Leger J., Lyonnet S.,

Pelet A., Czernichow P., Chatterjee K., Polak M.;
 "A novel loss-of-function mutation in TTF-2 is associated with
 congenital hypothyroidism, thyroid agenesis and cleft palate.";
 Hum. Mol. Genet. 11:2051-2059(2002).
 CC -|- FUNCTION: Probable transcription factor. Could be involved in
 CC thyroid gland organogenesis.
 CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -|- TISSUE SPECIFICITY: Detected in adult brain, placenta, lung,
 CC liver, skeletal muscle, kidney, pancreas, heart, colon, small
 CC intestine testis and thymus. Expression was strongest in heart and
 CC pancreas.
 CC -|- DISEASE: Defects in FOXE1 are the cause of Bamforth-Lazarus
 CC syndrome [MIM:241850]. A disease associated with thyroid agenesis,
 CC cleft palate and choanal atresia.
 CC -|- SIMILARITY: Contains 1 fork-head domain.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U89995; AAC51294.1; -;
 DR EMBL; Y13386; CAA73816.1; -;
 DR HSPG; Q63245; ZHFF.
 DR TRANSFAC; T02782; -;
 DR Genew; HGNC:3806; FOXE1.
 DR MIM; 602617; -;
 DR MIM; 241850; -;
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
 DR GO; GO:0003700; P:transcription factor activity; TAS.
 DR GO; GO:0007398; P:ectoderm development; TAS.
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
 DR GO; GO:0000122; P:negative regulation of transcription from P. . .; TAS.
 DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR PRODOM; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS50039; FORK_HEAD_3; 1.
 DR DNA-binding; Nuclear protein; Transcription regulation;
 KW Disease mutation.
 FT DOMAIN 7 12 POLY-PRO.
 FT DNA_BIND 52 143 FORK-HEAD.
 FT DOMAIN 164 182 POLY-ALA.
 FT VARIANT 57 57 S->N (in Bamforth-Lazarus syndrome;
 without choanal atresia).
 FT /FTID=VAR_016882.
 FT VARIANT 65 65 A->V (in Bamforth-Lazarus syndrome).
 FT /FTID=VAR_008857.
 FT CONFLICT 178 182 MISSING (IN REF. 2).
 SQ SEQUENCE 376 AA; 38289 MW; 766534A73231572C CRC64;
 Alignment Scores:
 Pred. No.: 2,69e-11 Length: 376
 Score: 436.50 Matches: 145
 Percent Similarity: 42.82% Conservative: 28
 Best Local Similarity: 35.89% Mismatches: 124
 Query Match: 19.94% Indels: 107
 DB: 1 Gaps: 17
 US-10-087-080-31 (1-1209) x FXE1_HUMAN (1-376)
 QY 196 AGTCGGAGCG 255
 Db 2 ThrAlaGluSerGlyProProPro-----ProGlnProGluValLeuAlaThr 17
 QY 256 GTGCGGAGGCG 315


```

Db      340 rpSerProGlyAlaSerProTyrLeuLysGlnPro----- 352
Qy      1108 CACCTGTACTGCCCCCTGGCGGCAGCCCTGGAGGGGGCCTT 1154
Db      353 --ProlaLeuthrProSerAsnProAlaAlaSerAlaGlyLeu 367

RESULT 15
FXCL_MOUSE STANDARD; PRT; 553 AA.
AC Q61572; O88409; Q61582;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forkhead box protein c1 [Forkhead-related protein FKHL7] (Forkhead-
DE related transcription factor 3) (FRAC-3) (Transcription factor FKHL-1)
DE [Mesoderm/mesenchyme forkhead 1] (MF-1).
GN FOXc1 OR FKHL7 OR FRAC3 OR FKHL1 OR MF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98297351; PubMed=9635428;
RA Kume T., Deng K.Y., Winfrey V., Gould D.B., Walter M.A., Hogan B.L.M.;
RT "The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic
RL mouse mutation congenital hydrocephalus.";
Cell 93:985-996(1998).
[2]
RN SEQUENCE OF 69-179 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kastner K.H., Lee K.H., Schoendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RT regulated.";

```



```

RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [3]
RP SEQUENCE OF 71-187 FROM N.A.
RX MEDLINE=93387221; PubMed=8375339;
RA Sasaki H., Hogan B.L.;
RT "Differential expression of multiple fork head related genes during
RT gastrulation and axial pattern formation in the mouse embryo.";
RL Development 118:47-59(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed in many embryonic tissues, including
CC prechordal mesenchyme, periocular mesenchyme, meninges,
CC endothelial cells and kidney.
CC -1- DEVELOPMENTAL STAGE: Expressed during embryogenesis.
CC -1- SIMILARITY: Contains 1 fork-head domain.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF045017; AAC24209.1; -.
DR EMBL; L10406; RAA03159.1; -.
DR EMBL; X71939; CAA50741.1; -.
DR PIR; I49674; I49674.
DR HSP; Q63245; 2HFH.
DR TRANSFAC; T02426; -.
DR MGD; MGI:1347466; Foxcl.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FOXHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK HEAD_1; 1.
DR PROSITE; PS00658; FORK HEAD_2; 1.
DR PROSITE; PS00039; FORK HEAD_3; 1.
DR DNA-binding; Nuclear protein; Transcription regulation.
KW DNA-BINDING; POLY-ALA.
FT DOMAIN 28 33 POLY-ALA.
FT DNA_BIND 77 168 FORK-HEAD.
FT DOMAIN 169 173 POLY-ARG.
FT DOMAIN 194 197 POLY-PRO.
FT DOMAIN 264 274 POLY-SER.
FT DOMAIN 375 386 POLY-SER.
FT DOMAIN 444 451 POLY-GLY.
FT DOMAIN 453 456 POLY-GLY.
FT DOMAIN 486 496 POLY-ALA.
FT DOMAIN 180 186 VKDKEEK -> KKEITFI (IN REF. 3).
FT CONFLICT 180 186
SQ SEQUENCE 553 AA; 56953 MW; 3CDD12F69CA4F217 CRC64;

Alignment Scores:
Pred. No.: 4.4e-11 Length: 553
Score: 430.50 Matches: 154
Percent Similarity: 43.14% Conservative: 44
Best Local Similarity: 33.55% Mismatches: 119
Query Match: 19.67% Indels: 142
DB: Gaps: 23

US-10-087-080-31 (1-1209) x FXCL_MOUSE (1-553)
QY 115 TCCTGGCTCAGATGGGACTGGCGCGCCAGCGCTCCGCGCGCGCGCGCGCGAGAT 174
Db 6 SerValSerProAsnSerLeuGlyValValProTyrLeuGly 20
QY 175 ACGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
Db 21 -----GlyGluGlnSerTyrArgAlaAlaAlaAlaAlaAlaGlyGly 36
QY 211 CCGGCGCGGAGGAGGAGGATCCCGGCGAGCGAGCTCTCCAGCGGTGGCGGCGGCG 270
Db 37 -----TyrThrAlaMetPro-----AlaProMetSerValTyrSerHisProAla 51

271 GAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 330
52 HisAla-----GluGlnTyrProGlySerMetAlaArgAlaTyrGly----- 65
331 AAGCCATATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
66 ---ProTyrThrProGlnProGlnProGlnProGlnProGlnProGlnProGln 84
373 CGCTCATCGCATCGCATCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
85 AlaLeuIleThrMetAlaIleGlnAsnAlaProAspLysLysIleThrLeuAsnGly 104
433 AACGAGTACCTCATGGCGCAAGTCCCTTTTCCGCGCGCGCGCGCGCGCGCGCG 492
105 TyrGlnPheIleMetAspArgPheProPheTyrArgAspIleGlnGlyTyrGln 124
493 TCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
125 SerIleArgHisAsnLeuSerLeuAsnGluCysPheValLysValProArgAsp 144
553 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
145 LysPro---GlyTyrGlySerTyrTyrThrLeuAspProAspSerTyrAsnMet 163
613 GACGGGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
164 AsnGlySerPheLeuArgArgArgArgPheLysLysLysAspAlaValLysAsp 183
646 -----AGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
184 GluGluLysGlyArgLeuHisLeuGlnGluProProProGlnAlaGlyArgGln 203
661 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
204 AlaProAlaPro-----ProGluGlnAlaGluGly-----SerAlaPro 218
721 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
218 OGlnProProValArgIleGlnAspIleLysThrGluAsnGlyThrCysProSer 238
760 GCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816
238 OProGlnProLeuSerProAlaAlaAlaLeuGlySerGlySerAlaAlaThrVal 258
817 GACGACATCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 876
258 sIleGluSerProAspSerSerSerSerSerSerSerSerSerSerSerSer 278
877 ACGCTTCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 918
278 rIeuProSerAlaArgProLeuSerLeuAspAlaAlaGluProAlaPro----- 296
919 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 935
296 OGlnPro-AlaProProHisSerGlnGlyPheSerValAspAsnIleMetThr 316
936 -----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 990
316 erLeuArgGlySerProGlnGlySerAlaAlaGluLeu----- 328
991 GCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1044
329 --GlySerGlyLeuLeuAlaSerAlaAlaAlaSerSerArgAlaGlyIleAla 348
1045 CTGCG-----ACCTCTCC----- 1058
348 euAlaLeuGlyAlaTyrSerProGlyGlnSerSerLeuTyrSerProCysSerGln 368
1059 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
368 erSerSerAlaGlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlySers 388
1081 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137

```

```

Db      388 e:SerAlaAlaGly-ThrGlyGlyAlaAlaThrTyHisCysAsnLeuGln---AlaMet 406
QY      1138 GCCCTGCAGCGCGCCTTACTCCGNCGTCTCTGGCCCGCACCTGCTGCTACCCG 1188
Db      407 SerLeuTyAlaAlaGlyGluArg-----GlyGlyHisLeuGlnGlyPro 421

```

Search completed: March 29, 2004, 12:01:15
 Job time : 30.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 29, 2004, 11:56:36 ; Search time 67 Seconds
(without alignments)

11386.911 Million cell updates/sec

Title: US-10-087-080-31

Perfect score: 2189
Sequence: 1 atgaagtggagggttcgt.....tgagacgcctcctagcttga 1209

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlsp
-Q=/cgn2_1/USPTO_spool_p/US10087080/runat_29032004_114526_6542/app_query.fasta_1.1351
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO US10087080 -CGCN_1_86 -runat_29032004_114526_6542 -NCPU=6 -ICPU=3
-NO WMAP -LARGSEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1597	73.0	400	11 Q9JJ18	Q9jj18 mus musculus

ID	Q9JJ18	PRELIMINARY;	PRT;	400 AA.
AC	Q9JJ18			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)			
DE	HFH1 (Forkhead box Q1).			
GN	FOXQ1 OR HFH1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129;			
RC	MEDLINE=21207087; PubMed=11309849;			
RA	Hong H.-K.; Noveroske J.K.; Headon D.J.; Liu T.; Sy M.S.;			
RA	Justice M.J.; Chakravarti A.;			
RT	"The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice.";			
RL	Genesis 29:163-171(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FVB/N-3; TISSUE=Breast tumor;			

ALIGNMENTS

RESULT 1

ID	Q9JJ18	PRELIMINARY;	PRT;	400 AA.
AC	Q9JJ18			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)			
DE	HFH1 (Forkhead box Q1).			
GN	FOXQ1 OR HFH1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129;			
RC	MEDLINE=21207087; PubMed=11309849;			
RA	Hong H.-K.; Noveroske J.K.; Headon D.J.; Liu T.; Sy M.S.;			
RA	Justice M.J.; Chakravarti A.;			
RT	"The winged helix/forkhead transcription factor Foxq1 regulates differentiation of hair in satin mice.";			
RL	Genesis 29:163-171(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FVB/N-3; TISSUE=Breast tumor;			

OS Rattus rattus (Black rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10117;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21207067; PubMed=11309849;
 RA Hong H.-K., Noveroske J.K., Heaton D.J., Liu T., Sy M.S.,
 RA Justice M.J., Chakravarti A.,
 RT "The winged helix/forkhead transcription factor Foxq1 regulates
 RT differentiation of hair in satin mice."
 RL Genesis 29:163-171 (2001).
 DR EMBL; AF153193; AAF73234.1; -.
 DR HSSP; Q63245; 2HFH.
 DR GO; GO:0005634; P:nucleus; IEA.
 DR GO; GO:0003700; P:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR GO; GO:0001766; TF:fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 SQ SEQUENCE 400 AA; 41444 MW; 99318026C8E2AD19 CRC64;

Alignment Scores:
 Pred. No.: 3,32e-70 Length: 400
 Score: 1534.00 Matches: 311
 Percent Similarity: 79.90% Conservative: 15
 Best Local Similarity: 76.23% Mismatches: 68
 Query Match: 70.08% Indels: 14
 DB: 11 Gaps: 7

US-10-087-080-31 (1-1209) x 09JLN7 (1-400)

QY 1 ATGAGTTGGAGTCTTCCTCGCGCGCCACGGGCAACGAGCGAGTCACTG 60
 Db 1 MetLeuLeuGluValPheAlaProArgAlaAlaHisGlyAspMetGlySerAspLeu 20
 QY 61 GAGGCGCGCGGCGGAGCGACGCGCGCTCCCGCTGTGCGGCGGGGAGAGCTCCCTG 120
 Db 21 GluGlyAlaGlySerSerAspValProSerProLeuSerAlaAlaGlyAspSerLeu 40
 QY 121 GCTCAGATGGGACTGCGCGGCC--AAGCGTCCGGCGGGCGCGCCAGAGATACG 177
 Db 41 GlySerAspGlyAspCysAlaAlaAsnSerProAlaAlaGlyArgGlyAlaValAspLeu 60
 QY 178 CAG--GGCGAGCGGGAACAGAGTGGGAGCGGCGCGCGCGCGCGATCCCG 234
 Db 61 GluGlyGlyGlyGlyGluArgAsnSerSerGlyGlyAlaSerThrGlnAsp----- 77
 QY 235 GCAGCAGTGTGTCAGCGGTGTGCGGAGGCGCG-----GAGCGCGGCGCGGGG 288
 Db 78 -----AspProGluValThrAspGlySerArgThrGlnAlaSerProValGly 93
 QY 289 CCAGCGCGGCGCGCGGAGCGCGGAGTGCAGCAGCAAGCCATATACGCGGCGG 348
 Db 94 ProCysAlaGlySerValGlyGlyGluGlyAlaArgSerLysProTyrThrArg 113
 QY 349 CCCAAGCCCCCTACTCGTACATCGCTCATCGCCATCGGCATCGGCGGCGG 408
 Db 114 ProLysProProTyrSerTyrIleAlaLeuIleAlaMetAlaIleArgAspSerAlaGly 133
 QY 409 GGGCGCTTACCTGGCGGAGTCAACGAGTACCTCATGGCAAGTTCCTTTTCCGC 468
 Db 134 GlyArgLeuThrLeuAlaGluIleAsnGluTyrLeuMetGlyLysPheProPheArg 153
 QY 469 GGCAGTACACGGGTGGCGCAACTCCGTGCGCCCAACACTTTCGCTCAACGACTGCTC 528
 Db 154 GlySerTyrThrGlyTrpArgAsnSerValArgHisAsnLeuSerLeuAsnAspCysPhe 173

QY 529 GTCAAGTGTCTGCGGACCCCTCGCGCGCTCGGGCAAGCAACTACTGATGCTCAAC 588
 Db 174 ValLysValLeuArgAspProSerArgProTrpGlyLysAspAsnTyrTrpMetLeuAsn 193
 QY 589 CCCAAGCAGGAGTACACCTTCGCGGACGGGCTTCGCGCGCGCGCGCGAGCGCTCAGC 648
 Db 194 ProAsnSerGluTyrThrPheAlaAspGlyValPheArgArgArgLysArgLeuSer 213
 QY 649 CACGCGCGCGGCTCCCGCGCGCGCTCGCGCGCGAGGAGCGCGCGCGCTCCCGCGC 708
 Db 214 HisArgThrThrValSerAlaSerGlyLeuArgProGluGluAlaProProGlyProAla 233
 QY 709 ---GCCCG 755
 Db 234 GlyThrProGlnProAlaProThrAlaGlySerSerProIleAlaArgSerProAlaArg 253
 QY 766 CAGGAGGAGCG 825
 Db 254 GlnGluGluGlySerSerProAlaSerLysPheSerSerSerPheAlaIleAspSerIle 273
 QY 826 CTGCGCAAGCCCTTCGCGAGCGCTCGCTCAGGAGCACGGCGCGCGCGCGCGCGCGCTCAG 885
 Db 274 LeuSerLysProPheArgSerArgArgAspGlyAspProAlaLeuGlyValGlnLeuPro 293
 QY 886 TGGGCG 945
 Db 294 TrpSerAlaAlaProCysProProLeuArgAlaTyrProAlaLeuLeuProAlaSerSer 313
 QY 946 TGCAGGCGCGCTGCTGCGCGTCTCGCGGTACGGCGCGCGCGCGCGCGCGCGCGCGCG 1005
 Db 314 GlyGlyAlaLeuLeuProLeuCysAlaTyrGlyAlaGlyGluProThrLeuLeuAlaSer 333
 QY 1006 CGGAGCGCGCGAGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1065
 Db 334 ArgGlyAlaGluValGlnProAlaAla---ProLeuLeuAlaProLeuSerThrAla 352
 QY 1066 GCCCGCGCGCAAGCACTCCGAGCGCGCG---GCCGCGCGCGCGCGCGCGCGCGCGCG 1122
 Db 353 AlaProAlaLysProPheArgGlyProGluThrAlaGlyAlaAlaHisLeuTyrCysPro 372
 QY 1123 CTGCGGTGCGCGAGCGCTTCGAGGGCGCTTGTAGTCGNGCGTCTGCGCGCGCGCGCGCG 1182
 Db 373 LeuArgLeuProThrAlaLeuGlnAlaAlaAlaAlaCysGlyProGlyProHisLeuSer 392
 QY 1183 TACCGGTGGAGCGCTCCTAGCT 1206
 Db 393 TyrArgValGluThrLeuLeuAla 400

RESULT 3
 O35392 PRELIMINARY; PRT; 492 AA.
 AC O35392;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Forkhead 2.
 GN FOXD2 OR MF2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Mesenchyme;
 RX MEDLINE=98148839; PubMed=9510020;
 WT Wu S.C.-Y., Grindley J., Winnier G.E., Hargrett L., Hogan B.L.M.;
 RA "Mouse Mesenchyme forkhead 2 (Mf2): expression, DNA binding and
 RT induction by sonic hedgehog during somitogenesis.";
 RL Mech. Dev. 70:3-13 (1998).
 DR EMBL; AF023915; AA881275.1; -.
 DR HSSP; Q63245; 2HFH.
 DR TRANSPAC; T02492; -.
 DR MGD; MGI:1347471; Foxd2.

DR Pfam; PF00250; Fork head; 1.
 DR PRINTS; PR00053; FOXHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK HEAD 1; 1.
 DR PROSITE; PS00658; FORK HEAD 2; 1.
 DR PROSITE; PS00039; FORK HEAD 3; 1.
 SQ SEQUENCE 443 AA; 46429 MW; 409563037EA5911F CRC64;

Alignment Scores:
 Pred. No.: 3,648-18 Length: 443
 Score: 506.00 Matches: 160
 Percent Similarity: 43.18% Conservative: 33
 Best Local Similarity: 35.79% Mismatches: 154
 Query Match: 23.12% Indels: 100
 DB: 13 Gaps: 20

US-10-087-080-31 (1-1209) x P79770 (1-443)

QY 64 GCGCGCGCGCGC-----ACGACGCGCGCTCCCGCTGTCGCGC 102
 DB 7 GlyGlyGlyGlyCysGlyIleMetSerGluArgSerProGluGluProLeuSerGlu 26
 QY 103 GCGGAGAC-----CACTCCCTGGCTCAGATGGGACTGCGCGCGCAAGCCGCC 153
 DB 27 ValGluAspAlaAspIleAspValValGlyProProGlnAsp---GlyAlaLysTyrSer 45
 QY 154 GCGGCGCGCGCGCGACAGATACGACGAGCGCGCAACAGTGCAGGAGCGCGCGC 213
 DB 46 GluAspGluGluAspAspAspAspGluGluAspGluGluGlyGlyPro 65
 QY 214 -----GGCGGAGGAGCGC 228
 DB 66 TrpGlySerProAlaAlaAspGlyGlyProProSerAlaHisGlyValProGluArg 85
 QY 229 ATCCGCGCAGAGCTCTCAGCGTGGTGGCGAGGCGCGCGCGCGCGCGCG 288
 DB 86 LeuSerProAlaGlyAlaArgSerProArgAlaProGlyProArgProGlyLysArgAla 105
 QY 289 CCAGCG 348
 DB 106 AlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyLysAsn---ProLeuVal----- 122
 QY 349 CCCAGCGCCCTACTCGTACATCGGCTCATCGCATCGCGCATCGCGACTCGCGCGCG 408
 DB 123 ---LysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleLeuGlnSerProLys 141
 QY 409 GCGCGCTTACGCTGCGGAGATCAACGAGTACCTCATGCGCAAGTTCCTCTTCGCG 468
 DB 142 LysArgLeuThrLeuSerGluIleCysGluPheIleSerGlyArgPheProTyrTyrArg 161
 QY 469 GCGAGCTACCGGCTGGCGCACTCCGTGGCGCAACCTTCGCTCAAGCTGCTTC 528
 DB 162 GluLysPheProAlaTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsnAspCysPhe 181
 QY 529 GTCAAGGTCTGCGGACCCCTCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 588
 DB 182 ValLysIleProArgGluProGlyAsnPro---GlyLysGlyAsnTyrTrpThrLeuAsp 200
 QY 589 CCCAAGCGAGTACACCTTCCGCGCGGCTCTCCGCGCGCGCGCGCGCGCGCGCG 648
 DB 201 ProGluSerAlaAspMetPheAspAsnGlySerPheLeuArgArgArgLysArgPheLys 220
 QY 649 ---CACGCGCGCGCGTCCCGC-----GGCGCGCGCGCTGCGCGCGC 684
 DB 221 ArgHisGluGlnProAspProProHisProGluLeuLeuLeuArgAlaGlyAlaArgSer 240
 QY 685 GAGGAGCGCGCGC----- 696
 DB 241 ArgArgLeuProAlaArgLeuArgTyrGlyProTyrGlyTyrAsnTyrGlyLeuGlnLeu 260
 QY 697 ---GGCTCCCG 753

DB 261 GlnGlyLeuProProGly---ProProAla-ProProProProProArg----- 275
 QY 754 TCGCGCGCGCGCGCGCGCGCGCGCGCA-----GCCCGCGCGCAAGTCTCTCCAGCTCC 807
 DB 276 -ArgArgArgLeuArgGlyAlaPheProPheSerAlaProHisCysProLeuProValGln 295
 QY 808 TTCGCCATCGACAGCATCTCGCCAAAGCCCTTCCGAGCGCGTCTCAGGGACACGGCC 867
 DB 295 yProProSerAlaAlaSerValPheSerAlaAlaSerGlyLeuProSer----- 311
 QY 868 CCGCGGACGAGCTTCAGTGGCGCGCGCGCTGCGCGCGCTGCGCGGTTCCCGCGC 927
 DB 312 -----PheLeuGlyGlyLeuAncyArgLysSerPheTyrHisProGln 327
 QY 928 CTCCTCCCGCGCGCGCTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 987
 DB 327 nLeuSerProThr-AlaLeu-----ProAlaAlaLeuLeuGlnThrLeuLysProAsp 345
 QY 988 CCGCGCGCGCGCGCGCGCGCGCGAGGCGCGCTGCGCGCGCGCGCGCGCGCGC 1037
 DB 345 roThrProAlaGlyThrGlyGlyThrAlaAlaAlaThrAsnProSerArgProSerP 365
 QY 1038 -----CCT 1062
 DB 365 heSerIleAspAsnIleIleGlyAlaValProProProProProSerThrAsnProSerA 385
 QY 1063 GC 1112
 DB 385 laAlaPro-----AlaAlaProTyrProSerGlyGlnAlaGlyProProAlaGlnLeu 403
 QY 1113 -----GTACTGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1158
 DB 403 euAlaValLeuSerProAlaLeuAlaProSerGlnProHisGlyGlyLeuAlaHisGlu 423
 QY 1159 CGNCGTCTGCGCGCGCA 1175
 DB 423 roLeuLeuGlnProAla 428

RESULT 5
 Q8R210 PRELIMINARY; PRT; 371 AA.
 ID Q8R210 AC Q8R210;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Forkhead box E1.
 GN FOXE1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA MEDLINE=22190894; Pubmed=12203737;
 RX Dathan N., Parlato R., Rosica A., De Felice M., Di Lauro R.;
 RT "Distribution of the ttf2/foxe1 gene product is consistent with an
 RT important role in the development of foregut endoderm, palate and
 RT hair";
 RL Dev. Dyn. 224:450-456(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Parlato R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ457047; CAD29716.2; .
 SQ SEQUENCE 371 AA; 37767 MW; 8AB92160F2DDFAFA CRC64;

Alignment Scores:
 Pred. No.: 3,47e-16 Length: 371
 Score: 467.00 Matches: 150
 Percent Similarity: 46.44% Conservative: 26
 Best Local Similarity: 39.58% Mismatches: 131

Query Match: 21.33% Indels: 72
DB: 11 Gaps: 16
US-10-087-080-31 (1-1209) x Q8R210 (1-371)
QY 217 GCGAGGAGGATCCCG-----GCAGCAGCTGCTCCACCGTGTGCGGAG 264
DB 3 AlaGluSerAlaProProProGlnProGluThrLeuAlaAlaVallysGluGlu 22
QY 265 GCGCGGAGCG 312
DB 23 ArgGlyGluAlaAlaAlaGlyValProAlaGluAlaAlaGlyArgGlyAla 42
QY 313 GCGAGGAGGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
DB 43 Gly---GlyArgArgGlyArgProLeuGlnArgGlyLysProProTyrSerTyr 61
QY 373 CGCTCATCCGATCCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
DB 62 AlaLeuIleAlaMetAlaIleAlaHisAlaProGluArgGluThrLeuGlyGly 81
QY 433 AACGAGTACCTGAGCAAGTCCCTTTTCCGCGCGCGCGCGCGCGCGCGCG 492
DB 82 TyrIysPheIleThrGluArgPheProPheTyrArgAspAsnProLysTyrGlnAsn 101
QY 493 TCCGTGCGCACACCTTCGCTCAACGATGCTTCTCAAGTGTGCGCGCGCGCG 552
DB 102 SerIleArgHisLeuThrLeuAsnAspCysPheLeuLysIleProArgGluAlaGly 121
QY 553 CGGCG 612
DB 122 ArgPro---GlyIysGlyIysTyrTrpAlaLeuAspProAsnAlaGluAspMetPheGlu 140
QY 613 GACGGGTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648
DB 141 SerGlySerPheLeuArgGlyArgPheLysArgSerAspLeuSerThrTyrPro 160
QY 649 -----CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
DB 161 AlaTyrMetHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 180
QY 676 CTGCG 705
DB 181 AlaValProAlaAlaArgProAlaTyrProGlyAlaValTyrAlaGlyTyrAlaPro 200
QY 706 ---GCCG 762
DB 201 LeuAlaAlaProProProValTyrTyrProAlaAlaSerPro-----GlyProCys 217
QY 763 CGC-----CAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
DB 218 ArgValPheGlyLeuValProGluArg-ProLeuSerProAspLeuGlyProAlaProSe 237
QY 811 GCCATGACGATCTTCGCGAGCCCTTCGCGAGCCGCTCGCTCAGGACACGGCCCC 870
DB 237 rAlaAlaGlySerCysAlaPheAlaAlaAlaAlaAlaGlyAlaAlaGlyThr-GlySerP 257
QY 871 GGGACGACGCT---TCAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
DB 257 heGlnProAlaValCysThrGlyAlaArgProValAsnProAlaAlaTyrAlaAlaAT 277
QY 925 GCGTCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 981
DB 277 rAlaAlaGlyProAspGlyAlaTyrProGlnGlyAlaSerSerAlaLeu-----292
QY 982 GCGGAGCG 1041
DB 293 --PheAlaAlaAlaAlaGlyArgLeuAlaGly-----ProAlaSerProP 307
QY 1042 CTGCTTGACCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
DB 307 roAla-----GlyGlyGlySerGlyGlyValGluAlaThrValAspPheTyrGlyArgT 325

QY 1083 -----CCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
DB 325 hrSerProGlyGlnPheGlyAlaAlaLeuGlyProCysTyrAsnProGlyGlyGlnLeuG 345
QY 1138 GCCCTGAGCGCGCGCTAGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1190
DB 345 lyaAlaGlyGlyGlyAlaTyrHisSerArgHisAlaThrAlaTyrProGly 362
RESULT 6
O08771 PRELIMINARY; PRT; 370 AA.
AC O08771; 1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE TTF-2 protein.
GN TTF-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RA Zannini M., Avantaggiato V., Biffali E., Arnone M.I., Sato K.,
RA Pischetola M., Taylor B.A., Phillips S.J., Simeone A., Di Leuro R.,
RT developing thyroid which is consistent with a role in controlling the
RT onset of differentiation."
RL EMBO J. 0:0-0(0).
DR EMBL; Y11321; CAA72174.1; -.
DR HSP; O63245; 2HFH.
DR TRANSFAC; T01133; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001766; TF_Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
SQ SEQUENCE 370 AA; 37784 MW; EE92C88CEE960F1F CRC64;
Alignment Scores:
Pred. No.: 2,67e-15 Length: 370
Score: 449.50 Matches: 149
Percent Similarity: 46.32% Conservative: 27
Best Local Similarity: 39.21% Mismatches: 129
Query Match: 20.53% Indels: 75
DB: 11 Gaps: 17
US-10-087-080-31 (1-1209) x O08771 (1-370)
QY 217 GCGAGGAGGATCCCG-----GCAGCAGCTGCTCCACCGTGTGCGGAG 264
DB 3 AlaGluSerAlaProProProGlnProGluAlaLeuAlaAlaVallysGluGlu 22
QY 265 GCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
DB 23 ArgGlyGluAlaAlaAlaGlyValProAlaGluAlaAlaGlyArgGlyAla 41
QY 313 GCGAGGAGGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
DB 42 Gly---GlyArgArgGlyArgProLeuGlnArgGlyLysProProTyrSerTyr 60
QY 373 CGCTCATCCGATCCGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432
DB 61 AlaLeuIleAlaMetAlaIleAlaHisAlaProGluArgGluThrLeuGlyGly 80
QY 433 AACGAGTACCTGAGCAAGTCCCTTTTCCGCGCGCGCGCGCGCGCGCGCG 492

```

Db      81  TyLysPheIleThrGluArgPheProPheTyArgAspAsnProLysLeuTyrPheGlnAsn 100
QY      493  TCGGTGGCCACCAACCTTCGCTCAACGACTCGTTCGTCAGGTGCTGGCGGACCCCTCG 552
Db      101  SerIleArgHisLeuThrLeuAsnAspCysPheLeuLysIleProArgGluAlaGly 120
QY      553  CGGCCCTGGGGCAAGCAACACTACTGATGCTCAACCCCAACAGGAGTACACCTTCGCC 612
Db      121  ArgPro---GlyLeuGlyAsnTyrTrpAlaLeuAspProAsnAlaGluAspMetPheGlu 139
QY      613  GACGGGGTCTTCGGCGCGCGCGCAAGCGCTCAGC----- 648
Db      140  SerGlySerPheLeuArgArgArgArgPheLysArgSerAspLeuSerThrTyrPro 159
QY      649  -----CACCGCGCGCGCTGCCCGCG-----CCCGGG 675
Db      160  AlaTyrMetHisAspAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaPheProGly 179
QY      676  CTGGCGCGCGAGGAGCGCGCGCTCCCG----- 705
Db      180  AlaValProAlaAlaArgProAlaTyrProGlyAlaValTyrAlaGlyTyrAlaProPro 199
QY      706  ---GCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
Db      200  LeuAlaAlaProProValTyrTyrProAlaAlaSerPro-----GlyProCys 216
QY      763  CGC-----CAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
Db      217  ArgValPheGlyLeuValProGluArg-ProLeuSerProAspLeuGlyProAlaProSe 236
QY      811  GCCATCGCAGCATCTCGCGCAAGCCTTCGCCGCGCGCTCGCGCGCGCGCGCGCGCGCG 870
Db      236  rAlaAlaGlyGlySerCysAlaPheAlaAlaAlaAlaGlyAlaProGlyThr-GlySerP 256
QY      871  GGGACGACGCT---TCAGTGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
Db      256  heGlnProAlaValCysThrGlyAlaArgProValAsnProAlaAlaTyrAlaAlaAla 276
QY      925  GGCCTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 981
Db      276  yAlaGlyProAspGlyAlaTyrProGlnGlyAlaSerSerAlaLeu----- 291
QY      982  GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1041
Db      292  --PheAlaAlaAlaAlaGlyArgLeuAlaGly-----Prot 303
QY      1042  CTGCTTGCACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1082
Db      303  hrSerProThrAlaGlyGlySerGlyGlyValGluAlaThrValAspPheTyrGlyA 323
QY      1083  -----CCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1134
Db      323  rgThrSerProGlyGlnPheGlyAlaAlaLeuGlyProCysTyrAsnProSerGlyGlnL 343
QY      1135  GCAGCCCTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1190
Db      343  euGlyAlaGlyGlyGlyAlaTyrHisSerArgHisAlaThrAlaTyrProGly 361
RESULT 7
Q86UP7  PRELIMINARY; PRT; 553 AA.
AC      Q86UP7;
DT      01-JUN-2003 (TrEMBLrel. 24, Created)
DT      01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Forkhead winged/helix transcription factor mutant 2.
GN      FOXCI.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RA      Konatireddy S., Chakrabarti S., Mandal A.K., Reddy A.B., Sampath S.,
RA      Panicker S.G., Balasubramanian D.;
RT      "Human FOXCI gene.";
RL      Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY228704; AAP15181.1; -.
DR      GO; GO:0005634; C:nucleus; IEA.
DR      GO; GO:0003700; C:transcription factor activity; IEA.
DR      GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro; IPR001766; TF_Fork_head.
DR      Pfam; PF00250; Fork_head; 1.
DR      PRINTS; PR00053; FORKHEAD.
DR      ProDom; PD000425; TF_Fork_head; 1.
DR      SMART; SM00339; FH_1.
DR      PROSITE; PS00657; FORK_HEAD_1; 1.
DR      PROSITE; PS00658; FORK_HEAD_2; 1.
DR      PROSITE; PS00309; FORK_HEAD_3; 1.
SQ      SEQUENCE 553 AA; 56784 MW; B2B7017111CE5125 CRC64;

Alignment Scores:
Pred. No.:      2,11e-14      Length:      553
Score:          431.50      Matches:    146
Percent Similarity: 42.05%      Conservative: 39
Best Local Similarity: 33.18%      Mismatches: 123
Query Match:      19.71%      Indels:     132
DB:              4          Gaps:       18

US-10-087-080-31 (1-1209) x Q86UP7 (1-553)
QY      214  GCGCGGAGAGGCGGATCCCGGAGAGCTGCTGTCAGCGGTGTGGCGGAGGCGCGGAG 273
Db      20  GlyGlyGlnSerTyrTyrArgAlaAlaAlaAlaAla-----AlaGlyGlyTyr 37
QY      274  GCGCGGCGCGGCGCGCGCGCGCG----- 297
Db      38  ThrAlaMetProAlaProMetSerValTyrSerHisProAlaHisAlaGluGlnTyrPro 57
QY      298  GCGCGCGCGGAGCGCGGAGGTGACGAGCAAGCATATACGCGCGCGCGCG----- 351
Db      58  GlyGlyMetAlaArgAlaTyrGly-----ProTyrThrProGlnProGlnPro 73
QY      352  -----AAGCCCCCTACTCGTACATCGCGTCTATCGCATCGCATCGCGCATCGCGGAC 399
Db      74  LysAspMetValLysProProTyrSerTyrIleAlaLeuIleThrMetAlaIleGlnAsn 93
QY      400  TCGCGCGCGCGCGCTTGACGCTGGCGGAGATCAACAGAGTACCTCATGCGCAAGTTCGCC 459
Db      94  AlaProAspLysLysIleThrLeuAsnGlyIleTyrGlnPheIleMetAspArgPhePro 113
QY      460  TTTTTCGCGCGGACACACGCGCTGGCGCAACTCCGTCGCGCCACACCTTCGCTCAAC 519
Db      114  PheTyrArgAspAsnLysGlnGlyTrpGlnAsnSerIleArgHisAsnLeuSerLeuAsn 133
QY      520  GACTGCTTCGTCAGGTGCTGGCGGACCCCTCGCGCGCTCGCGCGCTGGGCGCAAGCAACT 579
Db      134  GluCysPheValLysValProArgAspAspLysPro---GlyLysGlySerTyrTyr 152
QY      580  ATGCTCAACCCCAACAGCGAGTACACCTTCGCGGACGCGGTCTTCGCCCGCGCGCGCAAG 639
Db      153  ThrLeuAspProAspSerTyrAsnLysPheGluAsnGlySerPheLeuArgArgArg 172
QY      640  GCGCTCAGC-----CACCGCGCG 657
Db      173  ArgPheLysLysLysAspAlaValLysAspLysGluLysAspArgLeuHisLeuLys 192
QY      658  CCGGTCCCGCGCGCGCGCTGGCGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 708
Db      193  GluProProProGlyArgGlnProProProAlaProGluGlnAlaAspGlyAsn 212
QY      709  GCCCGCGCGCGCGCGCGCGCGCG-----CGGCTCGCGCGCGCG 747
Db      213  AlaPro-GlyProGlnProProValArgIleGlnAspIleLysThrGluAsnGlyTh 232

```


